

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

MILLENNIUM, L.P. and EON-NET, L.P.,

08-CV-02945 (WHP)(DFE)

Plaintiffs,

v.

PITNEY BOWES INC.,

Defendant.

DECLARATION OF PETER C. SCHECHTER

1. I am attorney duly admitted to practice in the States of New York and before this Court, and am a Partner with the firm of Edwards Angell Palmer & Dodge LLP, which represents defendant Pitney Bowes Inc. ("PBI"). I submit this declaration upon personal knowledge and in support of PBI's motion to dismiss with prejudice the complaint brought against it by plaintiffs Millennium, L.P. and Eon-Net, L.P. for failure to state a claim upon which relief can be granted pursuant to Rule 12(b)(6) of the Federal Rules of Civil Procedure, or in the alternative, for an order pursuant to Rule 12(e) of the Federal Rules of Civil Procedure requiring a more definite statement of plaintiffs' claims.

2. Annexed as **Exhibit A** hereto is a true and correct copy of United States Patent Number 5,258,855.

3. Annexed as **Exhibit B** hereto is a true and correct copy of United States Patent Number 5,369,508.

4. Annexed as **Exhibit C** hereto is a true and correct copy of United States Patent Number 5,625,465.

5. Annexed as **Exhibit D** hereto is a true and correct copy of United States Patent Number 5,768,416.

6. Annexed as **Exhibit E** hereto is a true and correct copy of United States Patent Number 6,094,505.

7. Annexed as **Exhibit F** hereto is a true and correct copy of United States Patent Number 6,683,697.

8. Annexed as **Exhibit G** hereto is a true and correct copy of United States Patent Number 7,075,673.

9. Annexed as **Exhibit H** hereto is a true and correct copy of United States Patent Number 7,184,162.

10. I declare under penalty of perjury that the foregoing is true and correct.

Dated: September 2, 2008
New York, N.Y.



PETER C. SCHECHTER

United States Patent [19]

Lech et al.

[11] Patent Number: 5,258,855

[45] Date of Patent: Nov. 2, 1993

[54] INFORMATION PROCESSING
METHODOLOGY

[75] Inventors: Robert Lech, Jackson; Mitchell A. Medina, Essex Fells; Catherine B. Elias, Plainsboro, all of N.J.

[73] Assignee: System X, L. P., New York, N.Y.

[21] Appl. No.: 672,865

[22] Filed: Mar. 20, 1991

[51] Int. Cl.⁵ H04H 1/40
 [52] U.S. Cl. 358/462; 358/448;
 358/453; 382/61; 382/48
 [58] Field of Search 358/400, 401, 403, 447,
 358/448, 449, 451, 452, 453, 460, 462, 463, 467,
 470, 471, 474; 382/61, 48

[56] References Cited

U.S. PATENT DOCUMENTS

4,034,343	7/1977	Wilmer	382/37
4,667,248	5/1987	Kanno	358/451
5,034,990	7/1991	Klees	382/50
5,095,445	3/1992	Sekiguchi	358/403
5,140,650	8/1992	Casey et al.	382/61
5,153,927	10/1992	Yamanari	382/61

OTHER PUBLICATIONS

Que's Computer User's Dictionary 2nd Ed. Bryan Pfaffenberger (author); ©1991; p. 144.

Primary Examiner—Edward L. Coles, Sr.

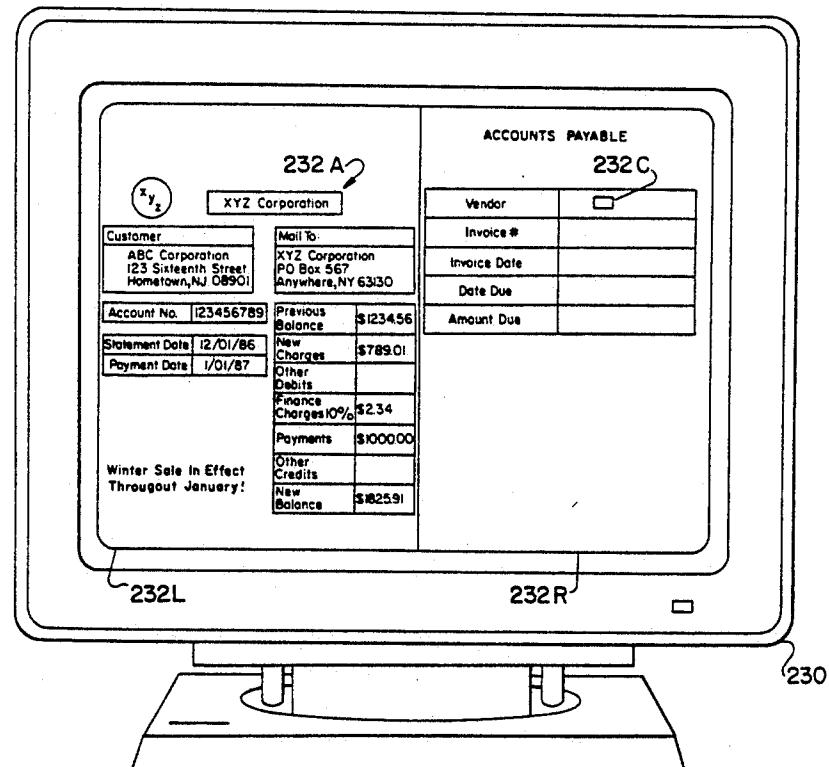
Assistant Examiner—Jerome Grant, II

Attorney, Agent, or Firm—Foley & Lardner

[57] ABSTRACT

An information processing methodology gives rise to an application program interface which includes an automated digitizing unit, such as a scanner, which inputs information from a diversity of hard copy documents and stores information from the hard copy documents into a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which designate portions of the stored document information required by a particular application program. The selected stored document information is then placed into the transmission format required by a particular application program in accordance with transmission format instructions. After the information has been transmission formatted, the information is transmitted to the application program. In one operational mode, the interface interactively prompts the user to identify, on a display, portions of the hard copy documents containing information used in application programs or for storage.

77 Claims, 15 Drawing Sheets



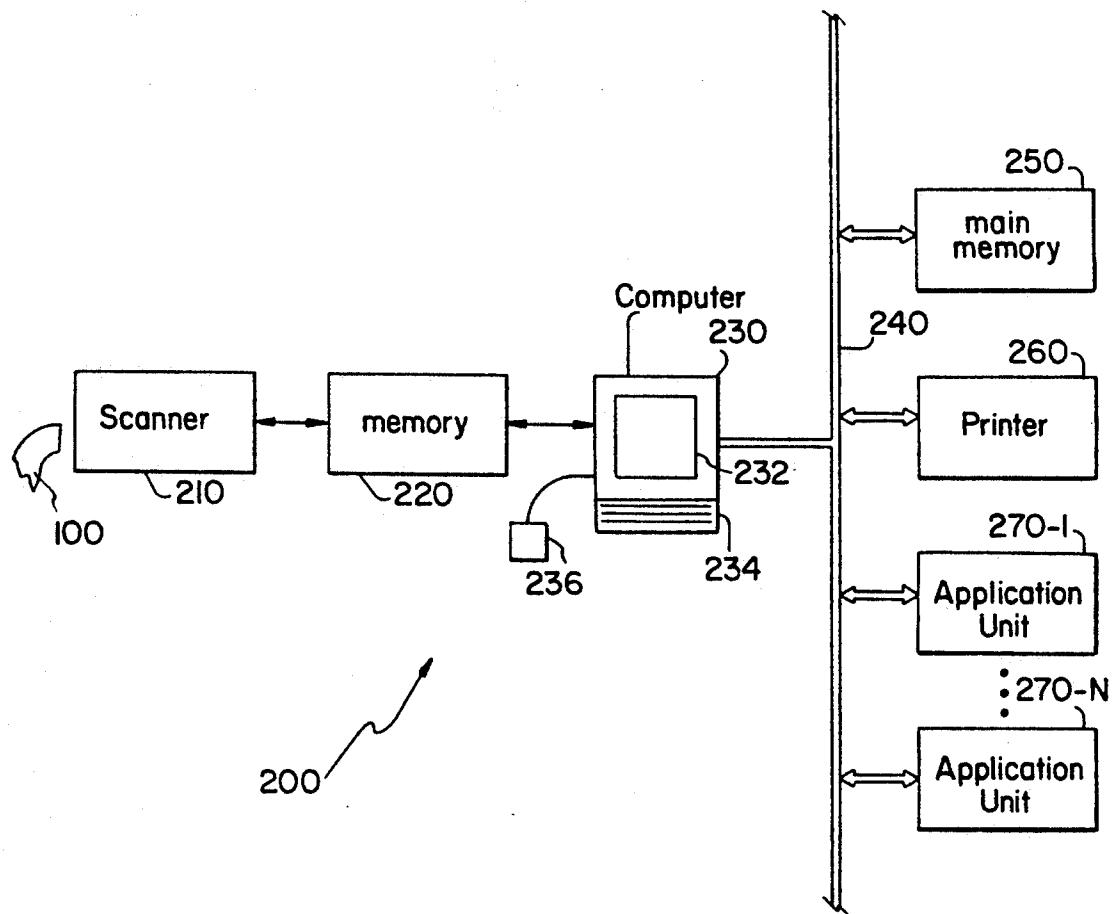
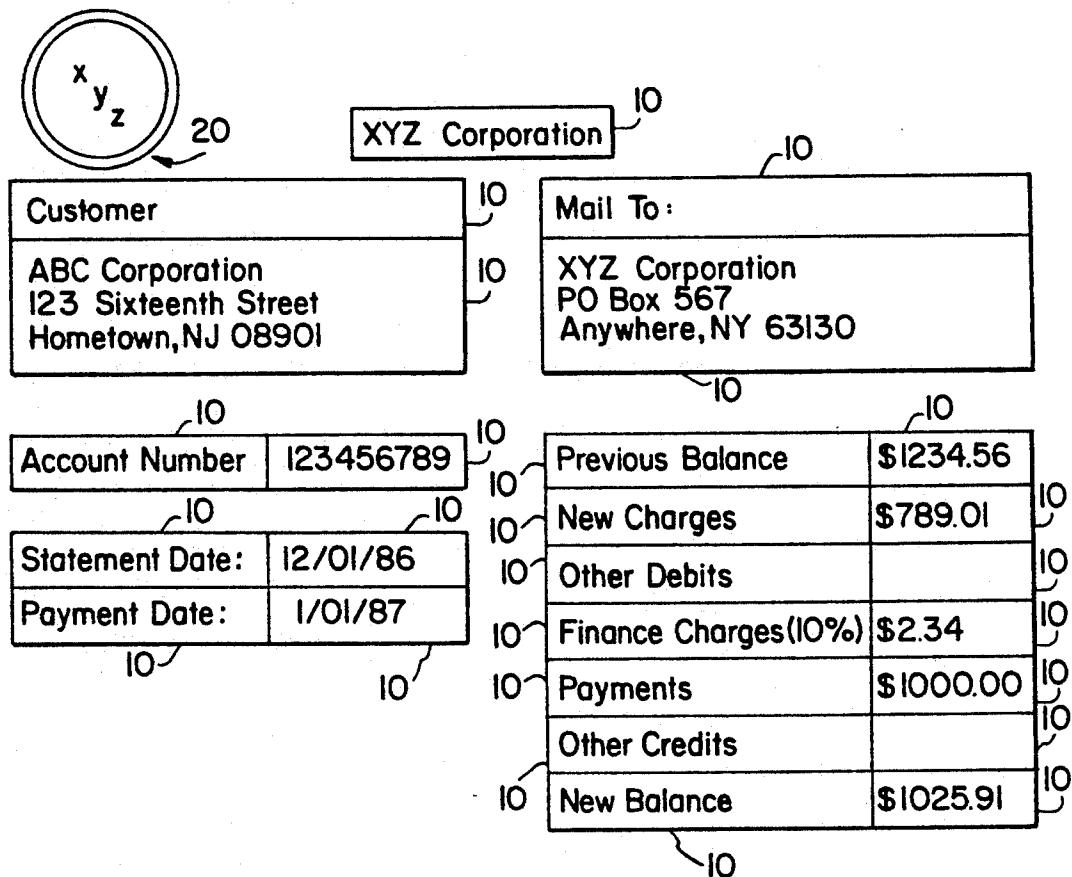


FIG. 1



Winter Sale In Effect Throughout January!

30

FIG. 2

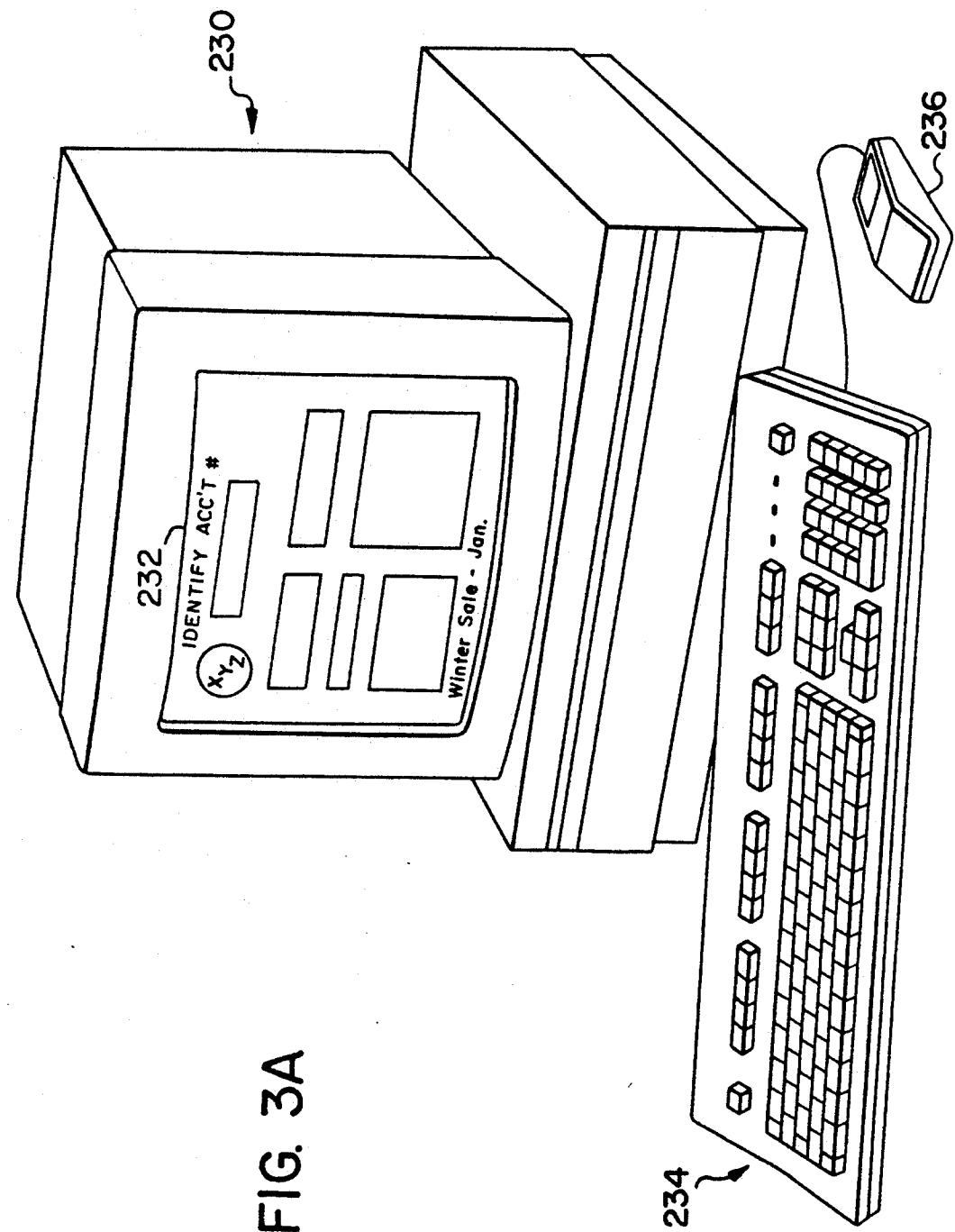


FIG. 3A

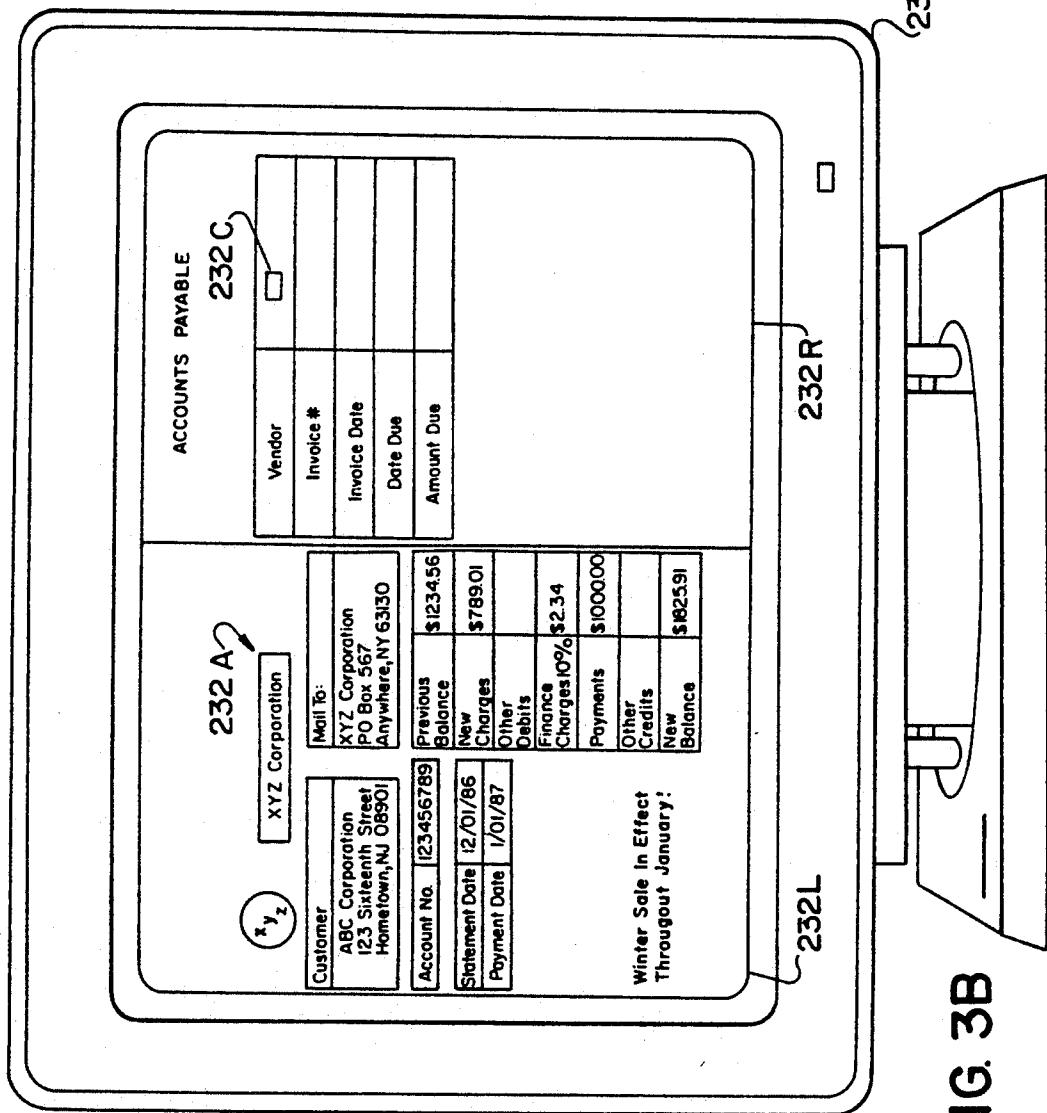


FIG. 3B

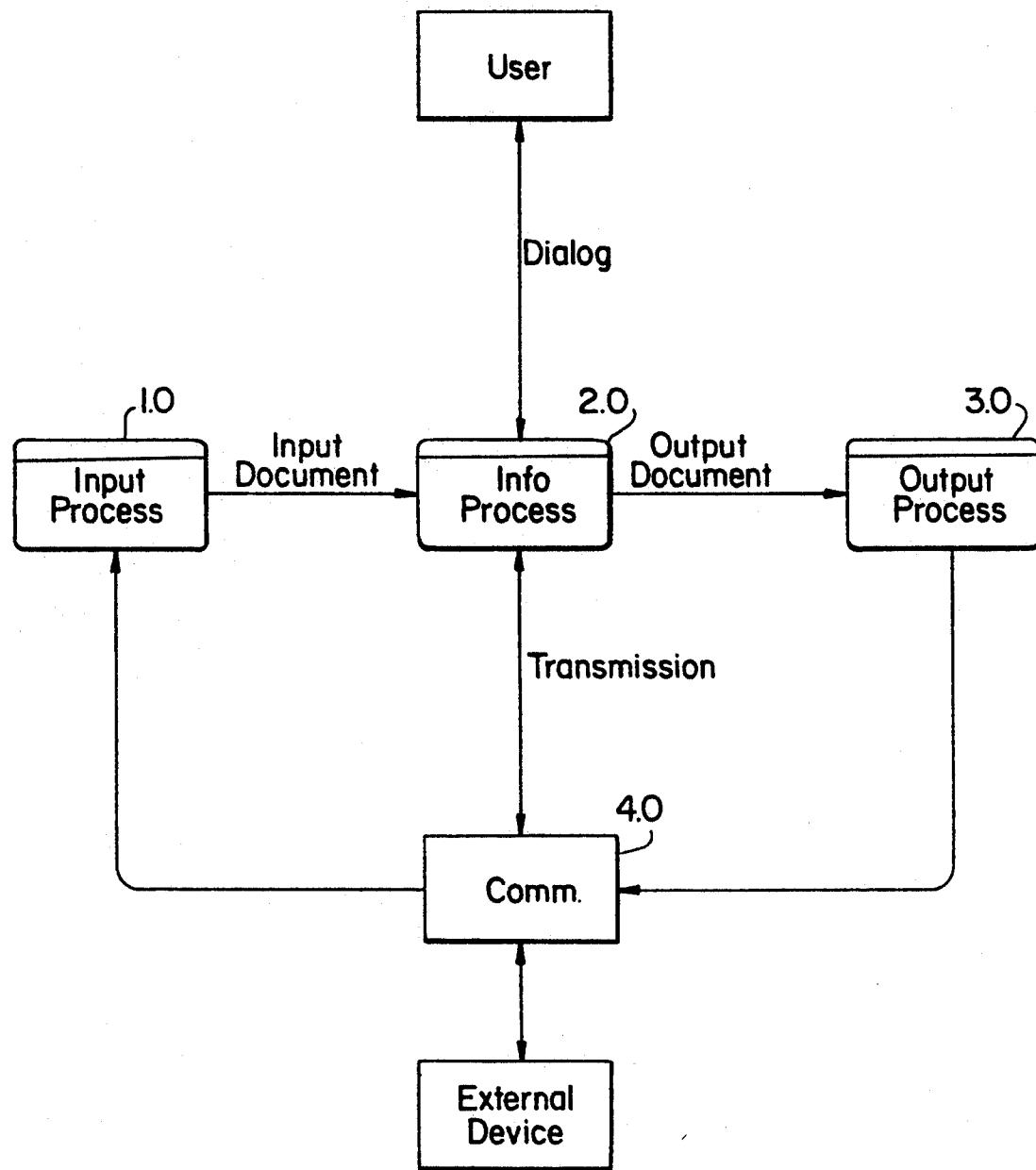


FIG. 4

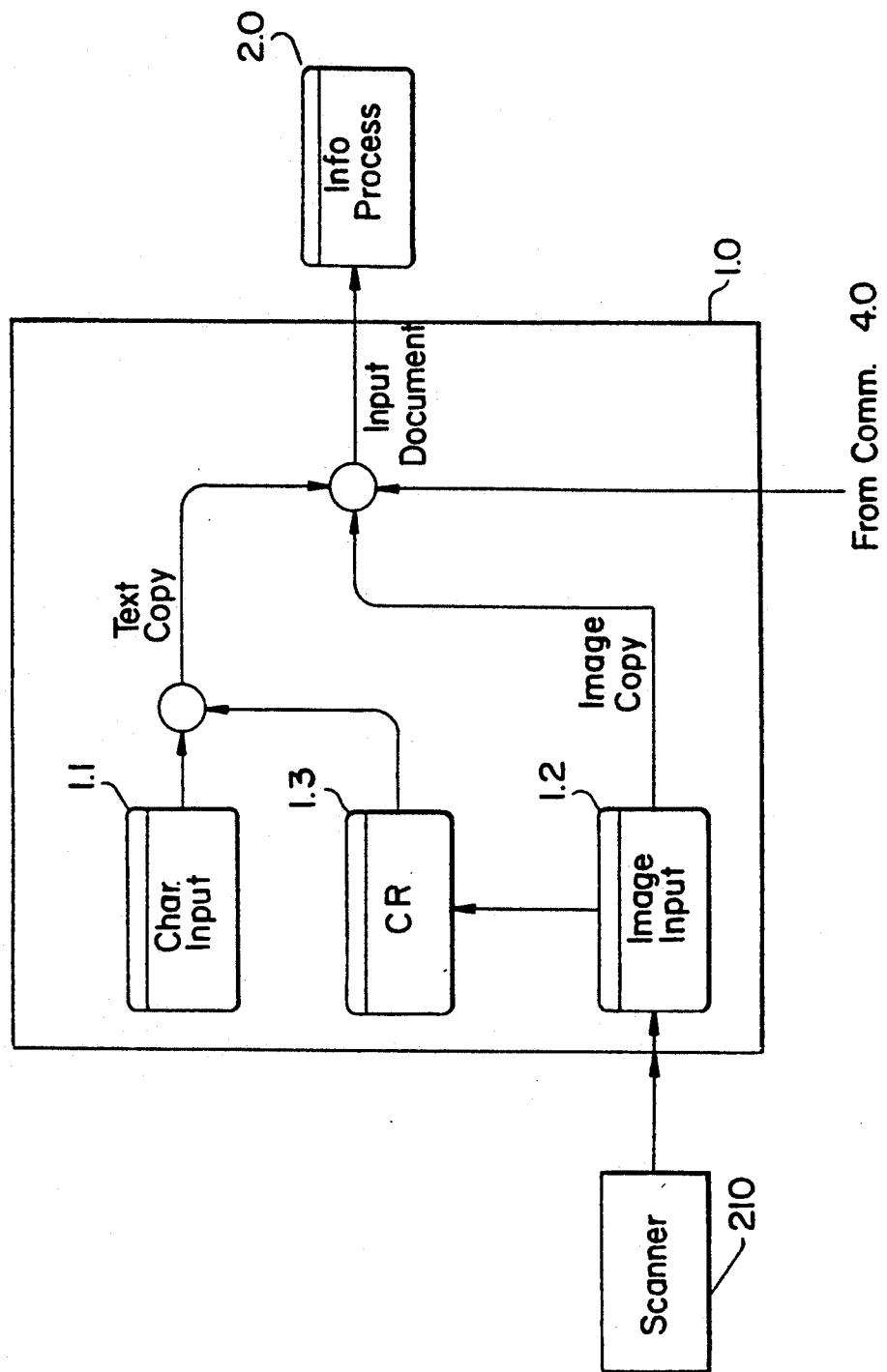
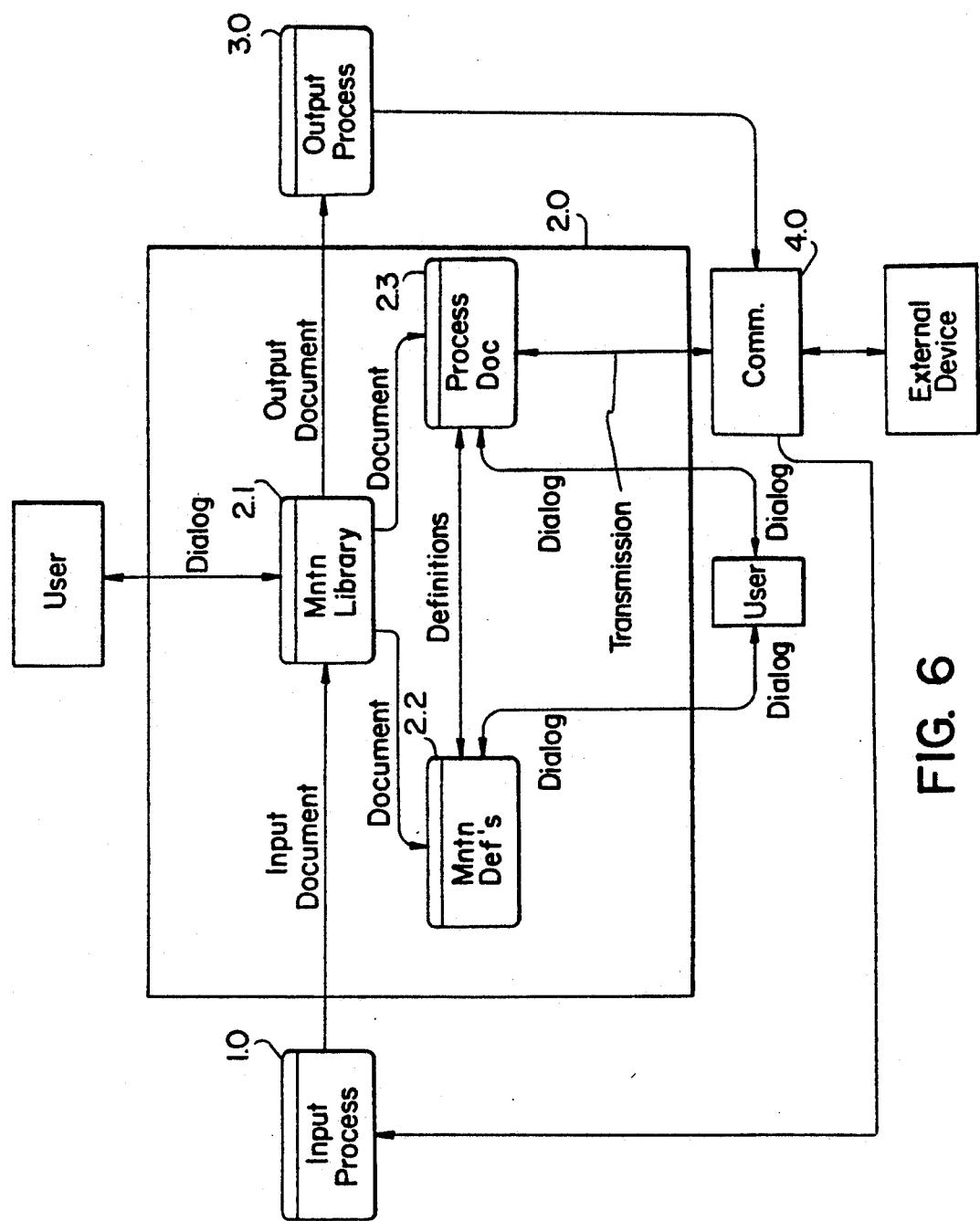


FIG. 5

**FIG. 6**

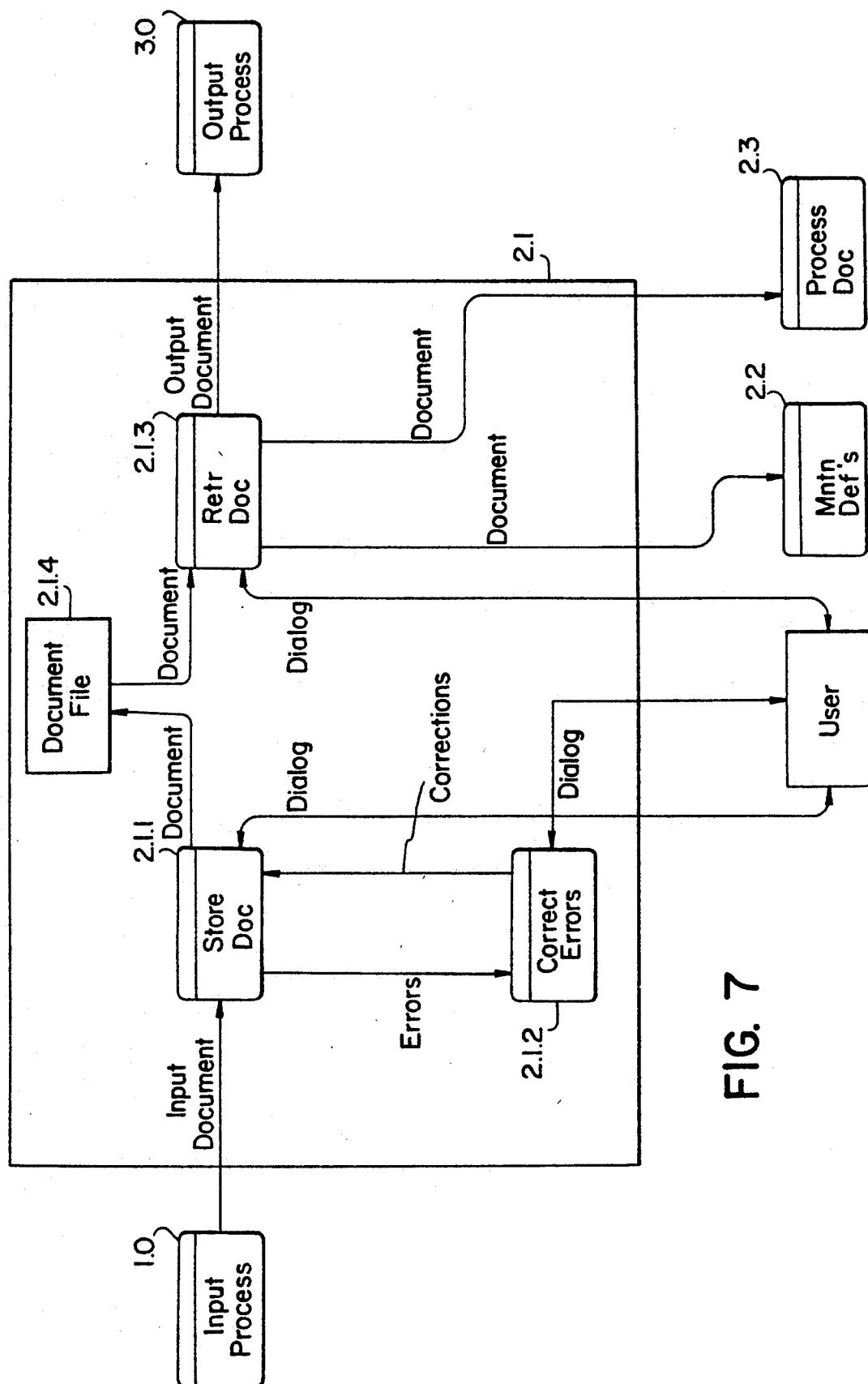


FIG. 7

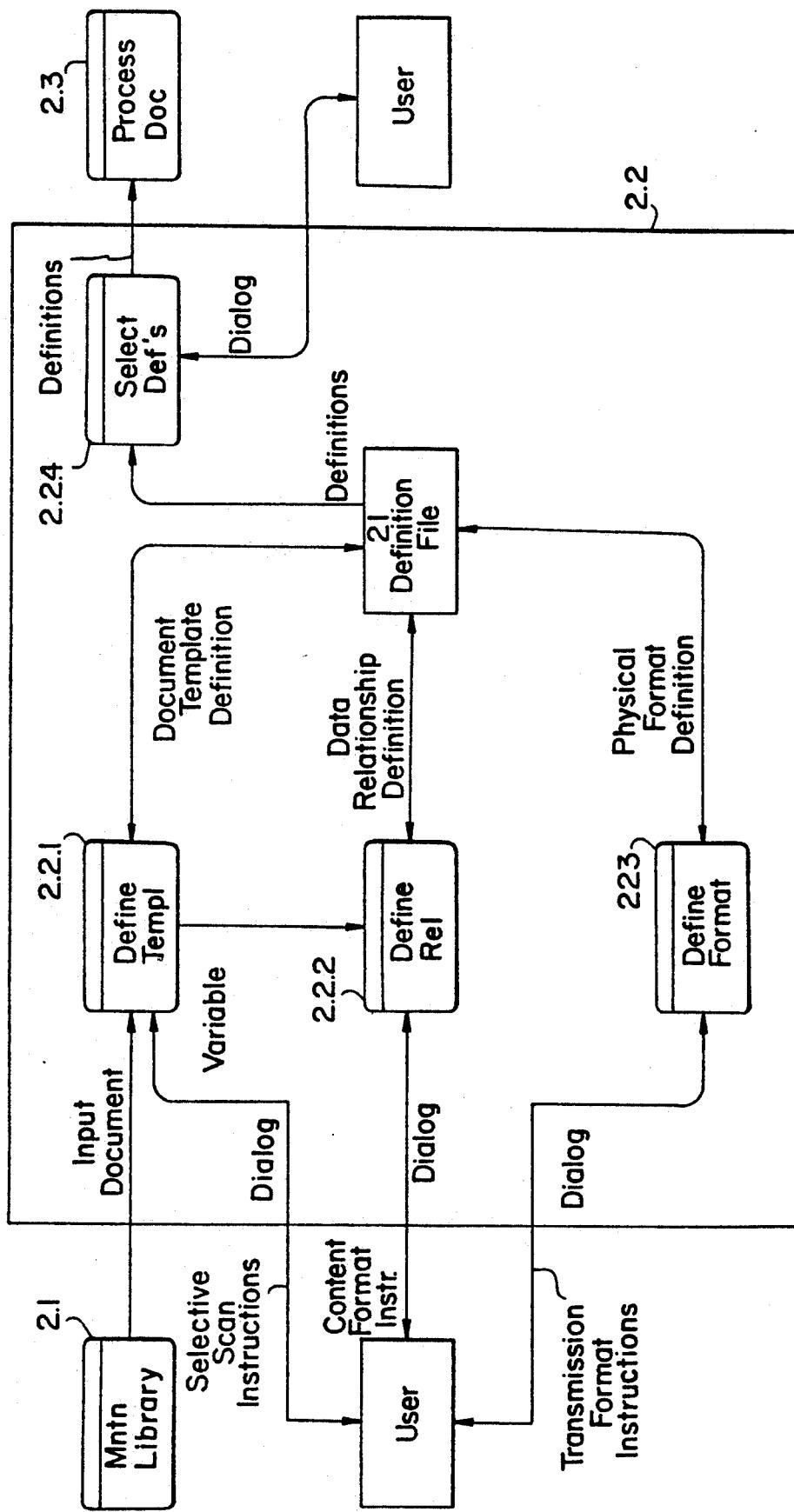


FIG. 8

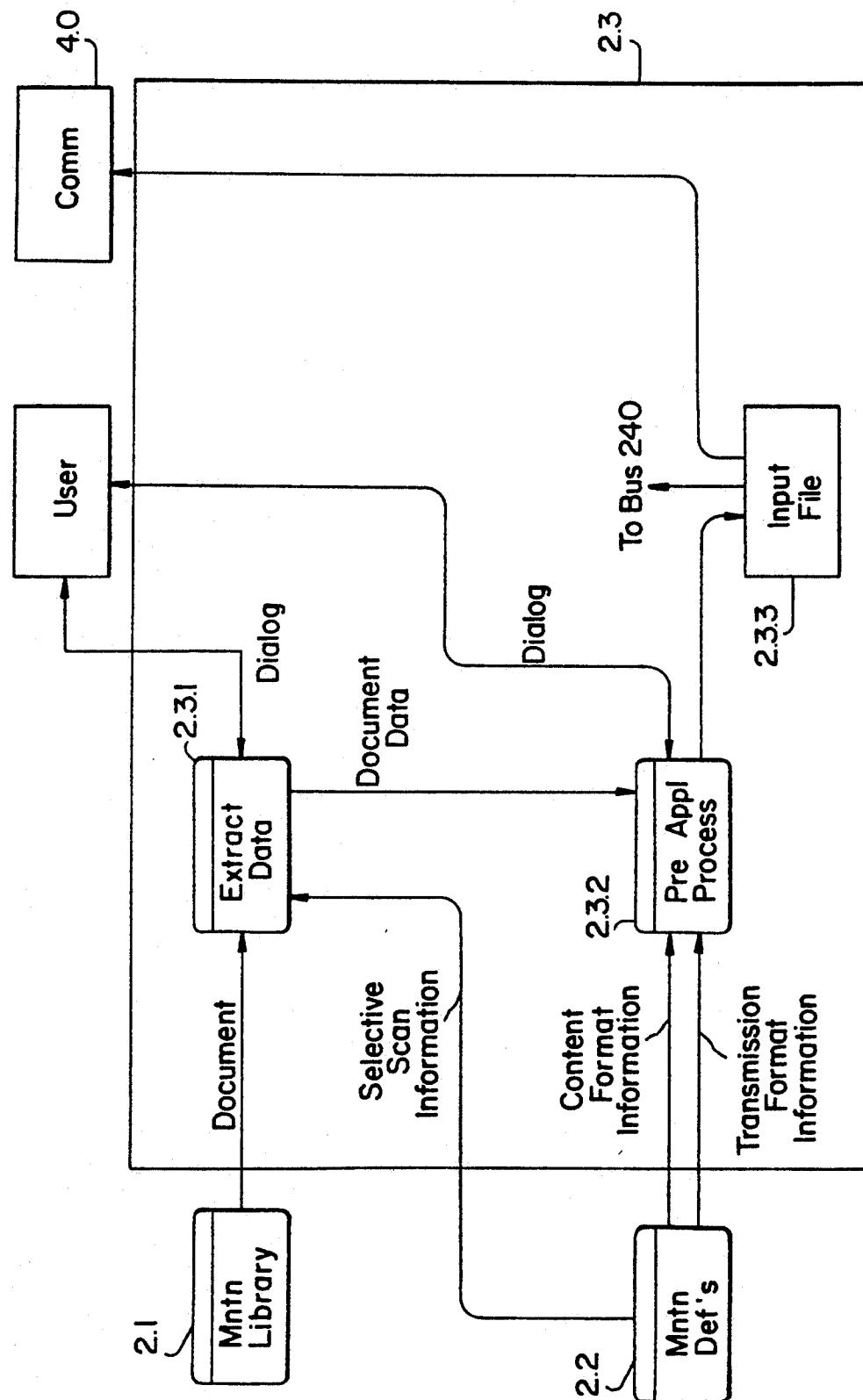


FIG. 9

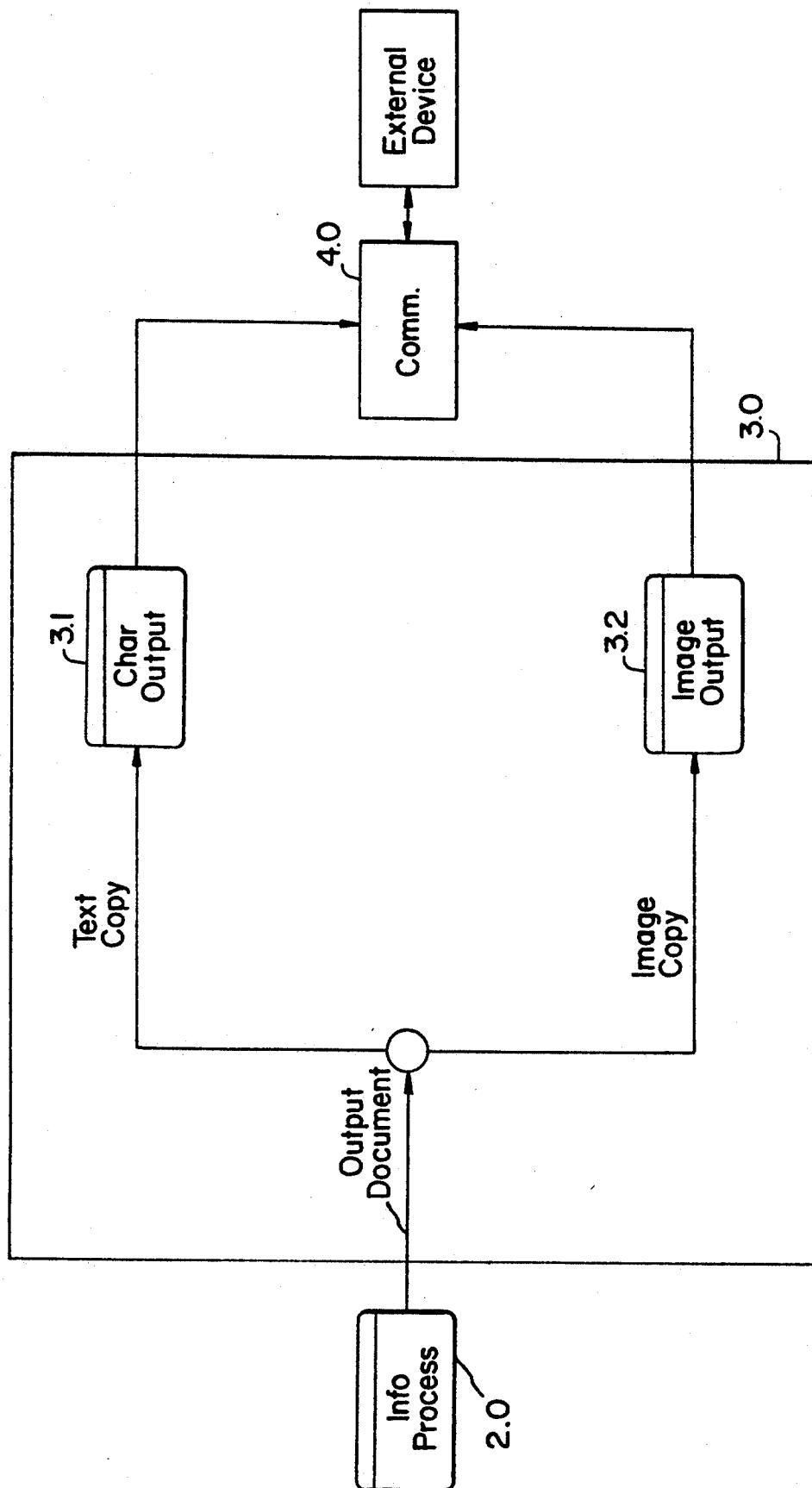


FIG. 10

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Heading 2	2
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
Other Credits	
New Balance	\$1025.91

FIG. II

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
New Balance	\$1025.91

FIG. 12A

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

FIG. 12B

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

FIG. 12C

```
>1>1>1>"XYZ Corporation"  
>2>2>25>+$123456789>  
>3>2>1>D12/01/86>  
>4>2>11>D12/15/86>  
>5>2>21>D01/01/87>  
>6>10>25>$1234.56  
>7>11>25>$789.01>  
>8>13>25>$2.34>  
>9>14>25>$1000.00>  
>10>16>25>$1025.91>
```

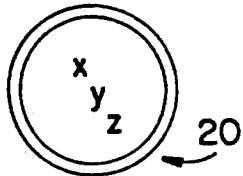
FIG. 13A

```
>1>1>1>$1234.56>  
>2>2>1>"XYZ Corporation"  
>3>3>1>"PO Box 567"  
>4>4>1>"Anywhere, NY 63130"
```

FIG. 13B

```
/1/1/1/$1234.56//  
/2/2/1/*XYZ Corporation*  
/3/3/1/*PO Box 567*  
/4/4/1/*Anywhere, NY 63130*
```

FIG. 13C



XYZ Corporation

Customer

ABC Corporation
123 Sixteenth Street
Hometown, NJ 08901

Mail To:

XYZ Corporation
PO Box 567
Anywhere NY, 63130

Account Number 123456789

Previous Balance \$1234.56

Statement Date 12/01/86

New Charges \$789.01

Payment Date 1/01/87

Other Debits

Finance Charges (10%) \$2.34

Payments \$1000.00

Other Credits

New Balance \$1025.91

Winter Sale in Effect Throughout January

30

FIG. 14

INFORMATION PROCESSING METHODOLOGY

BACKGROUND OF THE INVENTION

The invention is directed to a system for efficiently processing information originating from hard copy documents. More specifically, the invention is directed to a hard copy document to application program interface which minimizes the need to manually process hard copy documents.

In the past, information contained on hard copy documents was manually entered into a computer via the input controller of a particular computer. The original document was then filed away for future reference. Automatic input of data was limited to the input of Magnetic Ink Character Recognition (MICR) data and to Optical Character Recognition (OCR) data. This fixed-position data was forwarded directly to a dedicated computer application specifically designed to accommodate the input format. In more recent years, typewritten text has been mechanically inputted into a computer via a text file. Examples of this latter type of system are word processors and photo-typesetters.

These conventional systems have limitations which decrease the efficiency of processing information from a hard copy document. For example, the systems discussed above are limited in their application to MICR, OCR, or typewritten data. Parsing and processing data is limited to the particular requirements of the particular computer application which requires the input data. In addition, in these conventional systems, the actual hard copy document must be retained for future reference at great expense.

In a sophisticated computer network, different users may require different portions of the information contained on a hard copy document. For example, if the hard copy document is an invoice returned with payment of a bill, the accounting department may need all of the monetary information contained on the bill while the mailroom may need only customer address information, to update a customer's address. Therefore, there is a need for a system in which specific information from a hard copy document can be selectively distributed to various users.

Another problem with conventional systems is that users, even within the same company, may require that the information extracted from a hard copy document be transmitted to a particular application program in a specific transmission format. For example, one department in a company may use a particular application program which must receive information using a particular character as a delimiter and other departments may require the information in a different format using different delimiters.

Another problem, particularly for small businesses, is that current systems can not efficiently accommodate the inputting of information from a diversity of hard copy documents. A large business which receives many forms in the same format can afford a system which inputs a high volume of information in that format into memory. For example, it is cost-effective for a bank which processes hundreds of thousands of checks a month to buy a dedicated machine which can read information off of checks having a rigidly defined, or fixed, format. However, as the diversity of forms received by a business increases relative to the number of forms that must be processed, it becomes less cost-effective to design a dedicated machine for processing each

type of form format. This problem is particularly significant in small businesses which may, for example, receive fifty invoices a month, all in different, non-fixed, formats. It is frequently not cost-effective for a small business to design dedicated systems for inputting information in each of these various formats. This leaves a small business with no other practical alternative than to manually input the information off of each invoice each month.

10

SUMMARY OF THE INVENTION

It is an object of the invention, therefore, to provide an application program interface which allows a user to select specific portions of information extracted from a diversity of hard copy documents and allows the user to direct portions of this information to several different users in accordance with the needs of the particular user.

It is also an object of the invention to provide a cost-effective system for inputting hard copy documents which can accommodate hard copy documents in a diversity of formats.

It is another object of the invention to provide an application program interface which allows a user to put information, which is to be transmitted, into a particular transmission format, based upon the needs of the receiver of the information.

It is a further object of the invention to provide an application program interface which will allow the extraction, selection, formatting, routing, and storage of information from a hard copy document in a comprehensive manner such that the hard copy document itself need not be retained.

It is another object of the invention to provide a system which reduces the amount of manual labor required to process information originating from a hard copy document.

A further object of the invention is to reduce the time required to process information originating from a hard copy document so that a higher volume of transactions involving hard copy documents can be processed.

The invention provides an application program interface which inputs a diversity of hard copy documents using an automated digitizing unit and which stores information from the hard copy documents in a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which define portions of the stored document information required by a particular application unit. Selected stored document information is then formatted into the transmission format used by the particular application program based on transmission format instructions. The transmission formatted selected stored document information is then transmitted to the particular application program. The hard copy documents may contain textual information or image information or both.

The interface operates in three different modes.

In a first mode, the interface extracts all of the information from hard copy documents and stores this information in memory. Parsing of various portions of the extracted information is performed in accordance with content instructions.

In a second mode, the user operates interactively with the interface by use of a display and an input device, such as a mouse. In this second mode, a hard copy document is inputted and displayed on the display. The

30

35

40

45

50

55

60

65

interface then prompts the user to identify the location of various information. For example, the interface can ask the user to identify the location of address information on the hard copy document. In response, the user positions the mouse to identify address information using a cursor. The identified information is then stored as address information in memory. Subsequently, the interface again prompts the user to identify other pieces of information, which are then stored in the appropriate locations in memory. This process proceeds until all of the information which is desired to be extracted off of the hard copy document is stored in memory.

In a third mode of operation, selected portions of information are extracted off of hard copy documents in accordance with predetermined location information which has been specified by the user. For example, the user can define a template which specifies the location of information on hard copy documents. Templates can be formed in conjunction with second mode operation. Alternatively, the user can instruct the interface to search hard copy documents for a particular character or symbol, located on the hard copy documents. The information desired to be extracted off of the hard copy documents is specified relative to the location of this character or symbol.

The interface can also prompt or receive from an applications program or another information processing system, required information, content instructions, and format instructions.

Other objects, features, and advantages of the invention will be apparent from the following detailed description of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be described in further detail below with reference to the accompanying drawings, in which:

FIG. 1 illustrates hardware for implementing a preferred embodiment of the instant invention;

FIG. 2 illustrates an example of a hard copy document containing information to be processed by the instant invention;

FIGS. 3A and 3B are enlarged views of the computer of FIG. 1 used to explain how the invention interactively prompts a user to identify information;

FIG. 4 is an overall data flow diagram for the FIG. 1 preferred embodiment;

FIG. 5 is a detailed input data flow diagram for the FIG. 1 preferred embodiment;

FIG. 6 is a detailed information processing data flow diagram for the FIG. 1 preferred embodiment;

FIG. 7 is a more detailed information processing data flow diagram for the maintain library module of FIG. 6;

FIG. 8 is a more detailed information processing data flow diagram for the maintain definitions module of FIG. 6;

FIG. 9 is a more detailed information processing data flow diagram for the process document module of FIG. 6;

FIG. 10 is a detailed output data flow diagram for the FIG. 1 preferred embodiment;

FIG. 11 lists data corresponding to the hard copy document of FIG. 2;

FIGS. 12A, 12B, and 12C illustrate examples of data which can be selected from the extracted data of FIG. 11 in accordance with content instructions;

FIGS. 13A, 13B, and 13C illustrate examples of the data of FIGS. 12A, 12B, and 12C formatted in accor-

dance with various transmission format instructions to form input files; and

FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Hardware

The invention provides an interface between information originating from a hard copy document and a computer application unit which uses the information. The computer application unit can be a particular computer application program or a device which is controlled in accordance with instructions or information from the hard copy document. The invention also allows storing a copy of the hard copy document in a memory and retrieving the copy of the hard copy document. By providing a comprehensive and integrated system which can accommodate almost all of the possible uses of information contained on a hard copy document, the instant invention allows for a paperless office.

The invention includes hardware and software necessary to extract, retrieve, and process information from the hard copy document. A copy of the actual image of the hard copy document is stored in memory. Textual information extracted from the hard copy document is also stored in memory. Textual information is information, such as alphanumeric characters, which is recognized on the hard copy document and which is stored in a form which corresponds to the particular recognized character. For example, the extracted characters can be stored in the ASCII format in an electronic memory.

The user can have all of the information extracted from the hard copy document and stored in memory. Alternatively, the interface can interactively prompt the user to identify specific pieces of information for storage. The interface can also extract specific pieces of information using a predefined template. The interface can also prompt or receive from another information processing system or an applications program desired information, content instructions, and format instructions.

The instant invention also provides for parsing information extracted from the hard copy document and for directing this parsed information to specific users or application programs as an input file.

The invention also permits the user to define the transmission format of the input file for a particular computer application unit.

FIG. 1 illustrates hardware for implementing a preferred embodiment of a hard copy document to application program interface according to the instant invention. The interface 200 processes information extracted off of hard copy document 100 and provides information to application units 270 in a form required by each particular application unit. The interface extracts information off of a hard copy document 100 utilizing a scanner 210. The scanner 210 can be any type of scanner which extracts information off of hard copy documents, for example, an Optical Reader.

The scanned information is stored in a scanner memory 220 or in main memory 250, as will be described in greater detail below. If main memory 250 or another memory is available to store the scanned information, then scanner memory 220 can be omitted.

The information from scanner memory 220 or main memory 250 is transmitted to computer 230. In the preferred embodiment, computer 230 includes a display 232, a keyboard 234, and a mouse 236. The display 232 displays an image of the hard copy document itself and/or information necessary to process the information extracted off of the hard copy document.

The computer 230 is used to select portions of the stored document information contained in memory in accordance with content instructions which define portions of the stored document information required by an application unit. These content instructions may be provided by the application program. Alternatively, the content instructions can be inputted via an input device such as a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

The computer 230 is also used to format selected stored document information into the transmission format used by an application unit based on transmission format instructions. The transmission format instructions may be provided by the application program. Alternatively, the transmission format instructions can be inputted via a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

Thus, the computer 230 is used to generate an input file for a particular application unit. The computer 230 is connected to scanner memory 220, main, or permanent, memory 250, a printer 260, and application units 270, via a bus 240. Although FIG. 1 illustrates use of a bus to connect components together, it is understood that any routing or connecting link, implemented in hardware or software or both, can be employed instead of, or in addition to, a bus. Instructions to or in the computer 230 control the main memory 250, the printer 260, the application units 270, and the bus 240. Instructions to or in computer 230 can also control exchanges of information with scanner memory 220.

When the computer 230 generates an input file for a particular document, the computer 230 can send this input file directly to an application unit or can store this input file in the main memory 250 until required by an application unit. The main memory 250 may also optionally store a copy of the image information for the hard copy document and the textual information for the hard copy document. Thus, the image information and textual information from the hard copy document can be retrieved and printed out on printer 260. In addition, image and textual information stored in scanner memory 220 or in main memory 250 can be used to form additional input files at the time of input or at a later time, based on content instructions and transmission format instructions. Thus, the invention can, at the discretion of the user, eliminate the need to retain copies of hard copy documents, permitting a paperless office.

The application units 270 include particular application programs and devices which are controlled in accordance with information contained on hard copy document 100.

FIG. 2 illustrates an example of a hard copy document 100 which contains information to be processed by the instant invention. The document illustrated in FIG. 2 is a bill from XYZ Corporation to customer ABC Corporation. FIG. 2 is only an example of a type of document that can be processed by the instant invention.

In a first operational mode, the scanner 210 stores all of the information extracted off of hard copy document 100 in the scanner memory 220 or, alternatively, in main

memory 250. The extracted information is stored in two forms. The actual image of the hard copy document 100 is stored as image information in the scanner memory 220. In addition, the scanner memory 220 stores textual information recognized on the hard copy document 100 by, for example, employing standard character recognition software. In the preferred embodiment, the textual information is stored in ASCII format. The scanner memory 220 can be, for example, an electronic, magnetic, or optical memory.

FIG. 3A illustrates an enlarged view of the computer 230 of FIG. 1. This view will be used to describe a second mode of operation. In this second mode of operation, the hard copy document 100 is scanned and a copy of the document 100 is displayed on display 232 of computer 230, based on the contents of information temporarily stored in scanner memory 220. After the document is displayed on display 232, the computer 230 interactively prompts the user to identify the location of specific pieces of information on the hard copy document. In the FIG. 3A illustration, this prompt message is indicated as the message beginning with the arrow.

For example, the prompt message can ask the user to identify the location of account number information on the hard copy document. The user then uses an input device, such as keyboard 234 or mouse 236 or a touch screen, notepad, voice recognition device, or other input device to position a cursor on the display to identify the location of the information requested by the prompt message. For example, the cursor could be used to define a block (which could be highlighted) containing the requested information, followed by a mouse "enter" click. In this example, the user would move the mouse to identify the location of the account number information contained on the hard copy document 100. The computer 230 then stores the information which has been identified by the user as account number information in the appropriate address or subfile or as the appropriate variable or parameter in memory. The computer then prompts the user to identify the location of other information on the hard copy document, such as, statement date information. The process proceeds until all of the desired information has been stored into the appropriate locations in memory.

FIG. 3B illustrates a variation of the second mode for interactively prompting the user for information. In FIG. 3B, the display is split into two portions. A left-hand portion 232L displays the image of the hard copy document and a right-hand portion 232R displays the required application program information. For example, in FIG. 3B, portion 232R displays a spreadsheet used by an application program. While observing the split display, the user can input instructions to associate specific pieces of information on the hard copy document (for example, the vendor name indicated by the mouse arrow 232A) with particular subfiles in memory (for example, the vendor field next to which the cursor 232C appears), using a mouse or other input device(s) or both. The split display also allows the user to generate content format instructions while observing the information required for a particular application program on the right-hand portion.

These second modes of operation are efficient for small businesses which receive a small number of a wide variety of invoices, since the user does not necessarily have to store all of the information that appears on the hard copy document. A further advantage is that data input is quicker, easier, and more accurate than with

previous keyboard methodology. In addition, by specifying the location on the hard copy document of information, the user may optionally create a template, to be described in further detail below, for each different type of invoice. This template is stored for future use when another hard copy document in the same format is received.

More specifically, instructions from computer 230 can direct the scanner 210 and scanner memory 220, and/or main memory 250, to scan and/or store only specific portions of hard copy document 100. After the interactive prompts required to obtain information for a desired application program, the unused information stored in scanner memory 220 or 250 can be erased. Further, scanning of a second identical document can be limited to only those portions of the document which contain needed information.

More specifically, in FIG. 2, the lines 10 drawn around certain portions of the document represent the areas which the user has previously identified as the portions of a document to be extracted by the scanner 210 and stored in scanner memory 220 and/or main memory 250. Since the logo 20 and the message 30 have not been identified as an area to be scanned and stored, these areas are not scanned and stored in subsequent documents. Since the user has previously associated each of the areas 10 with a specific subfile of information, e.g., the account number, the scanned information is stored in memory locations corresponding to that subfile.

Data Processing

FIGS. 4-10 illustrate the flow of data in the FIG. 1 preferred embodiment. FIG. 4 illustrates the overall data flow for the FIG. 1 preferred embodiment. The preferred embodiment includes an input process module 1.0, an information processing module 2.0, and an output processing module 3.0. The information processing module 2.0 is equipped to receive instructions from and transmit information to a user. The information processing module 2.0 can also transmit to and receive information from a remote external device through communication interface 4.0. Input process module 1.0 and output processing module 3.0 can also access communication interface 4.0. A module is implemented in hardware, software, or a combination of hardware and software. The specific implementation for a particular business application depends upon a variety of factors, for example, the relative costs of hardware and software implemented systems, the frequency with which a user will want to expand or modify the system, and the like.

FIG. 5 is a more detailed diagram of the input process module 1.0 of FIG. 4. The input process module 1.0 includes a character input module 1.1, an image input module 1.2, and, in the preferred embodiment, a character recognition device 1.3. The character input module inputs textual information, such as alphanumeric characters, from an input device such as keyboard 234. The image input module 1.2 inputs image information, for example, a digitized image of the actual appearance of hard copy document 100. Textual information can include textual input from an input device such as keyboard 234 and textual information extracted from the document by character recognition device 1.3. Both types of information comprise an input document which is transmitted to information processing module 2.0. In the FIG. 1 preferred embodiment, the processing per-

formed by input process module 1.0 occurs in scanner memory 220, computer 230, and main memory 250.

FIG. 6 illustrates information processing data flow for the FIG. 1 preferred embodiment, that is, FIG. 6 illustrates data flow in the information processing module 2.0.

The information processing module 2.0 includes a maintain library module 2.1, to be described in further detail below in conjunction with FIG. 7, a maintain definitions module 2.2, to be described in further detail below in conjunction with FIG. 8, and a process document module 2.3 to be described in further detail below in conjunction with FIG. 9.

The information processing module 2.0 is the module which coordinates and drives the entire system. In the preferred embodiment, the information processing module 2.0 is implemented primarily by computer 230.

FIG. 7 illustrates information processing data flow in the maintain library module 2.1. The maintain library module 2.1 maintains a library of image information, for example, a digitized image representing the actual appearance of the hard copy document, and textual information of the hard copy documents for reference during processing. This library can be incorporated within scanner memory 220, main memory 250, or another independent memory, for example, a RAM disk. The maintain library module 2.1 includes a store document module 2.1.1, a correct errors module 2.1.2, a retrieve document module 2.1.3, and a document file 2.1.4. These modules operate collectively to store, retrieve, and correct document information.

The store document module 2.1.1, prior to routing the document to the document file 2.1.4, may provide information on recognition errors which may have occurred while inputting the document. For example, the store document module 2.1.1 identifies that a character contained on hard copy document 100 was not recognized. The store document module 2.1.1 also optionally causes a copy of the document and its parsing to be displayed on the display 232 for confirmation by the user. The user may utilize this opportunity to identify any errors in the displayed document and, in conjunction with the correct errors module 2.1.2, to revise the document's parsing, if necessary, prior to storage of the document in memory. The module 2.1.1 also provides a facility for the user to name a particular hard copy document for cataloging, storage, and retrieval purposes. After the document is named, the store document module 2.1.1 stores copies of the document in the document file 2.1.4.

The correct errors module 2.1.2 processes instructions from the user to correct errors identified by the store document module 2.1.1 and errors that have been spotted by the user during the confirmation process.

The retrieve document module 2.1.3 permits the user to retrieve a copy of a document previously stored in the document file 2.1.4. As described above, long-term storage is provided by main memory 250, if necessary.

FIG. 8 illustrates a more detailed information processing data flow diagram for the maintain definitions module 2.2 of FIG. 6. The maintain definitions module 2.2 allows the user to define system and document parameters and maintains the definitions of these system and document parameters. The maintain definitions module 2.2 includes a define template module 2.2.1 which allows the user to specify the location of information on the document. This information provided by the user defines a template which is used to extract

information off the document and to associate the extracted information with a particular variable or subfile. These templates are illustrated by boxes 10 in the FIG. 2 example of a hard copy document. The maintain definitions module 2.2 can also access templates previously defined by the user and stored in main memory 250. Templates can also be provided as part of software packages developed by program developers.

The maintain definitions module 2.2 also includes a define relationships module 2.2.2. The define relationships module 2.2.2 allows the user to define data relationships, or logical relationships, between pieces of information extracted from the hard copy document. These pieces of information are then used to generate an input file for a selected computer application unit. The user defines these relationships by content instructions. Alternatively, content instructions to define relationships can be provided by application software. If the user provides these content instructions, the content instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. Examples of content instructions, data, and logical relationships will be described in further detail in conjunction with FIGS. 11 and 12A, 12B, and 12C.

The maintain definitions module 2.2 also includes a define format module 2.2.3. The define format module 2.2.3 allows the user to define transmission formats for an input file which is then transmitted to a selected computer application unit. Selection of the transmission format of the input file is accomplished by the user through use of transmission format instructions. Alternatively, the applications software itself can generate its own transmission format instructions. When the user must specify transmission format instructions, the transmission format instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. A further description of various transmission formats will be provided below in conjunction with FIGS. 12A, 12B, 12C, 13A, 13B, and 13C.

A select definitions module 2.2.4 is also included in the maintain definitions module 2.2. The select definitions module 2.2.4 allows the user to store and select a set of definitions to be used for processing the document. The definitions identify pieces of information on the document by, for example, absolute location, variable location, or relative location, or by proximity to key words and/or symbols. These definitions are described in further detail below by way of an illustrative example.

FIG. 9 illustrates a more detailed information processing data flow diagram for the process document module 2.3. The process document module 2.3 processes the document after the document has been stored in the system. The process document module 2.3 gathers the appropriate information which has been stored, and creates input file(s) 2.3.3 for the selected application unit. The process document module 2.3 then transmits the input file(s) via bus 240 and/or communication interface 4.0 to an application unit 270, an output device such as printer 260, or to main memory 250.

The process document module 2.3 includes an extract data module 2.3.1. This module extracts data off of the document in accordance with the user's instructions, for example, the user-defined template, or through the interactive mode.

The process document module 2.3 also includes a preapplication process module 2.3.2 which gathers and associates information extracted from the document in accordance with content instructions. This module prompts the user for any additional information required to satisfy the relationships defined by the content instructions. The preapplication process module 2.3.2 also places the selected information into the transmission format defined by the transmission format instructions.

The preapplication process module 2.3.2 also generates the input file 2.3.3 for the selected application in accordance with the appropriate instructions. The input file 2.3.3 is then transmitted to bus 240 and/or communication interface 4.0 for transmission to a particular application unit 270.

FIG. 10 illustrates a detailed output data flow diagram for output module 3.0. Output module 3.0 outputs a textual and/or image copy of the document. In the FIG. 1 preferred embodiment, output module 3.0 is implemented by printer 260, associated software, and associated interface circuitry.

Operation

25 Examples of operation of a preferred embodiment will now be described.

The user enters the system by providing instructions to the information processing module 2.0. The user then instructs the information processing module 2.0 to conduct maintain library processing, maintain definitions processing, or process document processing.

If the user selects maintain library processing, the user then provides instructions to maintain or modify the document library through the maintain library module 2.1. For example, the user can direct the inputting and storage of a hard copy document 100 or can retrieve and output a document. The user requests inputting of a document through the store document module 2.1.1. The system then prompts the user to specify a storage location for the inputted document. The document is then read-in by the input process module 1.0. A textual copy and/or an image copy are stored into the document file 2.1.4. Errors which have occurred during inputting are identified and corrected by the correct errors module 2.1.2 and the user. The corrections are reflected in the document information stored in document file 2.1.4.

The retrieve document module 2.1.3 is used to retrieve and output a document. The system prompts the user to specify the storage location of a document and the type of document copy, for example, a textual or an image copy, to be outputted. The document is then outputted by the output process module 3.0.

If the user initially selected maintain definitions processing, the user would instruct the system to maintain and/or modify parameter definitions through the maintain definitions module 2.2. For example, the user can define and maintain a document template for extracting selected portions of information off of the hard copy document. The user can use the template to extract selected portions of information off of the hard copy document when the document is originally inputted, or alternatively, the user can use the template to identify selected portions of information for extraction off of an image copy of the document. In creating the template, the user identifies pieces of information on the document to be extracted and assigns a variable name, or subfile, to each piece of data.

The location of data to be extracted can be defined in a number of ways other than by use of a template. For example, the user can designate the absolute location of information on the document with respect to a grid overlaid on the document, e.g., always on line 3, starting in column i. The user can also identify information by specifying the relative location of information to be extracted, e.g., always two lines below the piece of data named "salutation", starting in column 3. The user can also specify the location of information to be extracted by variable location specification. For example, if the hard copy document is a letter, the module would conduct a key word search for the term "Dear Sir:". Whenever this term "Dear Sir:" is located, this piece of data would be associated with the variable specified by the user, for example, the variable "salutation." In addition, a defined set of conventional symbols can be used to signify certain recurring data items for the convenience of users of the instant invention. For example, a "@" symbol can be used to delineate the vendor name as follows: "@XYZ Corporation@". Other examples of the use of symbols to delineate information will be described with reference to FIG. 14.

The maintains definition module 2.2 is also used to maintain data relationships in accordance with content instructions and to maintain input file formats in accordance with transmission format instructions. Relationships are defined and maintained between pieces of data, specified by, for example, the names of variables, through the define relationships module 2.2.2. The names of pieces of data on the document are retrieved by, for example, the define template module 2.2.1, and are passed to the define relationships module 2.2.2. The user may then provide any additional pieces of data needed to generate an input file for a particular application program or unit, such as an input file line number. The user, the applications software, and/or instructions previously stored in memory then establishes the contents of the input file by defining relationships between pieces of data using content instructions. Specific examples of content instructions will be discussed below in conjunction with FIGS. 11, 12A, 12B, 12C, 13A, 13B, and 13C.

The user and/or the applications software defines and maintains the transmission format of the input file to be used by a particular application program or unit through the define format module 2.2.3 in accordance with transmission format instructions. This is accomplished by defining the parameters to be used by the preapplication process module 2.3.2 in generating an input file. Parameters which would typically be required to generate an input file would include the character type, e.g., text or pixel; delimiters used between pieces of data, e.g., a slash or a semicolon; end of line characters, e.g., a carriage return or a line feed; and end of file characters. Examples of transmission formats will be described in further detail below in conjunction with FIGS. 11, 12A, 12B, 12C, 13A, 13B, and 13C.

If the user initially selected process document processing, the interface will then proceed to process the document through use of the process document module 2.3. For example, the user can extract specific portions of data from an image copy of a document, can generate an input file for transmission to an application program, or can directly process information interactively with an application program.

If the user desires to extract specific portions of data from an image copy of a hard copy document which has

already been stored in memory, the user uses the extract data module 2.3.1 to identify a document to be processed. The document is then retrieved by the retrieve document module 2.1.3 and passed to the extract data module 2.3.1. The user can also select parameter definitions through the select definitions module 2.2.4.

The selected document template or parameter definition is passed to the extract data module 2.3.1. The extract data module 2.3.1 extracts pieces of data from 10 the image copy of the document, as defined by the document template definition or the parameter definitions or both. This document data is then passed to preapplication process module 2.3.2.

The interface generates input file(s) 2.3.3 by use of the 15 preapplication process module 2.3.2. The selected data relationship definition, as defined by the content instructions, and the selected record format definitions, as defined by the transmission format instructions, are passed to the preapplication process module 2.3.2. The 20 preapplication process module 2.3.2 assembles the input file in accordance with the content instructions. The preapplication process module 2.3.2 also prompts the user for any additional pieces of data which need to be provided by the user. The input file is converted to the 25 desired transmission format in accordance with the transmission format instructions. This physically formatted data is then stored in the input file 2.3.3.

The user can also use an application program to process information by loading the particular application program into the computer 230 rather than by sending the input file to a remote application unit 270.

An illustrative example of the processing described above will now be described.

The user inputs instructions via keyboard 234 or another input device which indicate that the user desires to input and store a document. The computer 230 then prompts the user for the name of the document. In this example, the user desires to input the document of FIG. 2 and therefore names the document "XYZ Corp. Bill 30 Dec. 1, 1986." The computer then prompts the user to feed the hard copy document 100 into the scanner 210. The image of the hard copy document is displayed on display 232. The computer then prompts the user to identify the account number on the document. By use of 35 the mouse 236 or other input device to position a cursor on the display, the user indicates the location of the account number. The account number is then read-in to a subfile named "Account Number." This process proceeds until all of the desired information has been read-in and stored.

In this particular example, no errors were encountered while inputting the document. The user then directs that the document be stored for future reference in a document file.

Some time later, the user desires to retrieve and output the document and to generate input files based on information from the document. The computer 230 prompts the user for the name of the document and the type of output. The user responds with "XYZ Corp. Bill Dec. 1, 1986" for a printed textual copy. The document is then retrieved from the document file and passed to the printer 260 for printing.

In order to generate an input file for a specific application program, the user selects the option to define a document template for use when each month's XYZ Corporation bill arrives. Accordingly, the user instructs the system to display a copy of an XYZ Corporation bill on the display 232. The user then identifies pieces of

data by absolute locations. That is, the user assigns specific names to information located at specific portions of the document. In this example, the user would input the following information:

Vendor-text, line 1, one line, column 1, 80 characters;
Account number-numeric, line 6, one line, column 25, 9 characters;

Statement date-date, line 9, one line, column 25, 8 characters;

Payment date-date, line 11, one line, column 25, 8 characters;

Previous balance-currency, line 7, one line, column 75, 9 characters;

New charges-currency, line 8, one line, column 75, 9 characters;

Other debits-currency, line 10, one line, column 75, 9 characters;

Finance charges-currency, line 12, one line, column 75, 9 characters;

Payments-currency, line 13, one line, column 75, 9 characters;

Other credits-currency, line 14, one line, column 75, 9 characters;

New balance-currency, line 15, one line, column 75, 9 characters.

The user also identifies data with variable locations. In this particular example, a variable location is specified as follows:

Heading 2-line, value="Mail To:"

The identification of Heading 2 as line information means that the system will search for occurrences of the character string "Mail To:" and assign the line number which contains this character string to Heading 2.

The user also identifies data by relative locations. In this example, the user identifies the following relative location:

Mail To-text, Heading 2+1, 3 lines, column 60, 25 characters per line.

The instructions above instruct the system to assign the textual information beginning on one line after Heading 2 and continuing for 3 lines, in column 60, to the Mail To subfile.

As an alternative to inputting the actual line, column, and character numbers, the user can identify desired portions of the document by blocking, or highlighting, the desired portions using the mouse or other input device. In this case, the computer converts the highlighted portions into corresponding line, column, and character numbers.

FIG. 11 lists data corresponding to the hard copy document of FIG. 2 and the associated variable or subfile names.

Next, the user desires to define data relationships in accordance with content instructions. Examples of the type of contents which can be specified by a user are illustrated in FIGS. 12A, 12B, and 12C.

In this particular example, three separate departments of ABC Corporation require information from the XYZ Corporation bill. The first department requires vendor, account number, statement date, payment date, previous balance, new charges, debits, finance charges, payments, and new balance information. The second and third departments require mail to information and previous balance information. Each of these departments have their own application program which utilizes this information.

The user employs content instructions to designate how pieces of information, which have been extracted

off of hard copy document 100, are directed to particular departments, that is, particular application programs. FIG. 12A illustrates the contents of the information to be transmitted to the first department. FIG. 12B illustrates the information to be transmitted to the second department. FIG. 12C illustrates the information to be transmitted to the third department. The content instructions, therefore, parse the information shown in FIG. 11 to various application programs, as shown by FIGS. 12A, 12B, and 12C. Content instructions can also be used to identify additional pieces of data which are required for the input files of the particular application programs. In this particular example, the specific application programs from the three departments all require numeric record number information, numeric horizontal position information, numeric vertical position information, and date received information. The horizontal and vertical position information is used by the application program to specify the location of the received information on a spreadsheet application program, in this example. The user may know in advance the content format required by each application program, that is, in this example, the location and type of information specified on the spreadsheet. The user may also employ the split display mode described with reference to FIG. 3B to generate content format instructions.

Using the content instructions, the user establishes the following contents for the input file corresponding to FIG. 12A:

Record number, horizontal position, vertical position, vendor;

Record number, horizontal position, vertical position, account number;

Record number, horizontal position, vertical position, statement date;

Record number, horizontal position, vertical position, date received;

Record number, horizontal position, vertical position, payment date;

Record number, horizontal position, vertical position, previous balance;

Record number, horizontal position, vertical position, new charges;

Record number, horizontal position, vertical position, finance charges;

Record number, horizontal position, vertical position, payments;

Record number, horizontal position, vertical position, new balance.

Next, transmission format instructions are employed to define the transmission format of the input file for a specific application program or unit. FIG. 13A illustrates the transmission input file corresponding to FIG. 12A. FIG. 13B illustrates the transmission input file corresponding to FIG. 12B. FIG. 13C illustrates the transmission input file corresponding to FIG. 12C. A comparison of FIGS. 12B and 12C reveals that FIGS. 12B and 12C have the same contents. However, the information illustrated in FIG. 12B is being sent to a different application program than the information in FIG. 12C. These application programs require different transmission input formats, as illustrated in FIGS. 13B and 13C. More specifically, the application program that receives the input file illustrated in FIG. 13B uses the greater than sign as a delimiter whereas the application program which receives the transmission input file shown in FIG. 13C uses a back-slash as the delimiter.

After the contents and the transmission format for the input file have been defined, and any additional information has been inputted, the input file is assembled and transmitted to the particular application program.

FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention. The hard copy document illustrated in FIG. 14 is first scanned and information from the hard copy document is stored into a memory. The interface 200 then identifies portions of the hard copy document corresponding to various variables by recognizing a defined set of symbols. In the FIG. 14 example, triangles delineate the mailing address, circles delineate the statement date, and squares delineate the new charges. Information from these portions of the hard copy document is stored in the corresponding memory locations or subfiles for each variable. The same set of symbols can be used to identify the same information from one document to the next. Thus, even if the physical formats of documents are not fixed from one document to the next, a diversity of hard copy documents can be processed without manually inputting data by recognition of the defined symbols.

Examples of readily available application programs are Quicken and Lotus 1, 2, 3 both of which are widely utilized in the business community. Quicken, for example is an easy-to-utilize program for writing checks and preparing business records. Payee, amount and address information may readily be transmitted from scanner memory 220 and/or main memory 250 to the Quicken application program for check writing functions and ledger keeping purposes. Lotus is a well known spreadsheet program which may process data input into specified cells once this data is placed in conventional Lotus format.

Thus, the instant invention provides an integrated and comprehensive system for handling information from a hard copy document, thus permitting a paperless office. In addition, the invention permits data, extracted off of a hard copy document, to be easily manipulated into various logical and transmission formats required by a particular application unit. The invention also provides a low cost system for inputting information from a wide variety of hard copy documents into a memory.

The foregoing description has been set forth merely to illustrate preferred embodiments of the invention and is not intended to be limiting. Modifications are possible without departing from the scope of the invention.

For example, letters, checks, forms, pictures, reports, music scores, film, and other types of hard copy documents can be processed by the invention for accounts payable/receivable accounting, inventory control, record keeping, budgeting, data base management, music transcription, forms processing, computerized art, survey and questionnaire processing, statistical data analysis, correspondence processing and other applications.

Other automated digitizing units can be used in addition to or as an alternative to use of the scanner 210 as an input unit. Any electrical, magnetic, or optical device which extracts information off of a hard copy document, thereby eliminating the need to manually input significant amounts of information from the hard copy document is suitable for use as an automated digitizing unit. In addition, information can be input by user responses and digital and analog signals generated from various devices, and from computer files from other computer systems. Suitable hardware for inputting data

includes a keyboard, a light pen, a mouse, a touch screen, a laser scanner, a microphone, a tablet, a disk drive, a magnetic tape drive, and a modem.

The interface 200 can also output information in forms other than a hard copy of textual or image information. For example, the interface 200 can output system responses, computer files, and digital and analog signals for transmission to other computer systems or to control systems. Suitable hardware for outputting information includes a disk drive, a magnetic tape drive, a cathode ray tube, a plasma screen, a printer, a plotter, a film developer, an amplifier, and a modem.

Since modifications of the described embodiments incorporating the spirit and substance of the invention may occur to persons skilled in the art, the scope of the invention should be limited solely with respect to the appended claims and equivalents.

What is claimed is;

1. A method of processing information from a diversity of types of hard copy documents, said method comprising the steps of:

- (a) receiving output representing a diversity of types of hard copy documents from an automated digitizing unit and storing information from said diversity of types of hard copy documents into a memory, said information not fixed from one document to the next, said receiving step not preceded by scanning, via said automated digitizing unit, of a separate document containing format requirements;
- (b) recognizing portions of said hard copy documents corresponding to a first data field; and
- (c) storing information from said portions of said hard copy documents corresponding to said first data field into memory locations for said first data field.

2. A method as set forth in claim 1, wherein step (b) includes prompting identification of said portions of said hard copy documents corresponding to said first data field.

3. A method as set forth in claim 1, wherein step (c) includes storing textual information from said portions of said hard copy documents corresponding to said first data field into said memory locations for said first data field.

4. A method as set forth in claim 1, further comprising the steps of detecting and correcting errors resulting from automated digitizing.

5. A method as set forth in claim 1, further comprising the steps of utilizing a template to associate portions of said hard copy documents with specific data fields.

6. A method as set forth in claim 1, further comprising receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a location on said hard copy documents containing data corresponding to a specific data field.

7. A method as set forth in claim 1, further comprising receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a relative location on said hard copy documents containing data corresponding to a specific data field.

8. A method as set forth in claim 1, wherein step (b) includes displaying an image of a hard copy document on a display based on a contents of said memory.

9. A method as set forth in claim 8, further comprising the steps of:

- recognizing portions of said hard copy documents corresponding to a second data field, said portions of said hard copy documents corresponding to said

second data field being different from said portions of said hard copy documents corresponding to said first data field; and
storing information from said portions of said hard copy documents corresponding to said second data field into memory locations or said second data field.

10. A method as set forth in claim 1, further comprising the steps of:

recognizing portions of said hard copy documents corresponding to a second data field, said portions of said hard copy documents corresponding to said second data field being different from said portions of said hard copy documents corresponding to said first data field; and
storing information from said portions of said hard copy documents corresponding to said second data field into memory locations for said second data field.

11. A method as set forth in claim 10, further comprising the step of storing image information from said portions of said hard copy documents corresponding to said second data field into said memory locations for said second data field.

12. A method as set forth in claim 10, further comprising the step of storing textual information from said portions of said hard copy documents corresponding to said second data field into said memory locations for said second data field.

13. A method as set forth in claim 10, further comprising the step of prompting identification of said portions of said hard copy documents corresponding to said second data field.

14. A method of processing information from a diversity of types of hard copy documents, said method comprising the steps of:

(a) scanning a diversity of types of had copy documents and storing information from said diversity of types of hard copy documents into a memory, said information not fixed from one document to the next, said scanning not preceded b scanning of a separate document containing format requirements;

(b) recognizing portions of said hard copy documents corresponding to a first data field; and

(c) storing information from said portions of said hard copy documents corresponding to said first data field into memory locations for said first data field.

15. A method as set forth in claim 14, further comprising the steps of detecting and correcting errors resulting from said scanning.

16. A method as set forth in claim 14, further comprising receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a relative location on said hard copy documents containing data corresponding to a specific data field.

17. A method as set forth in claim 14, further comprising receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a location on said hard copy documents containing data corresponding to a specific data field.

18. A method as set forth in claim 14, further comprising the step of utilizing a template to associate portions of said hard copy documents with specific data fields.

19. A Method as set forth in claim 14, wherein step (c) includes storing textual information from said portions of said hard copy documents corresponding to said first data field into said memory locations for said first data field.

20. A method as set forth in claim 14, wherein step (b) includes prompting identification of said portions of said hard copy documents corresponding to said first data field.

21. A method as set forth in claim 14, wherein step (b) includes displaying an image of a hard copy document on a display based on a contents of said memory.

22. A method as set forth in claim 21, further comprising the steps of:

15 recognizing portions of said hard copy documents corresponding to a second data field, said portions of said hard copy documents corresponding to said second data field being different from said portions of said hard copy documents corresponding to said first data field; and

storing information from said portions of said hard copy documents corresponding to said second data field into memory locations for said second data field.

23. A method as set forth in claim 14, further comprising the steps of:

recognizing portions of said hard copy documents corresponding to a second data field, said portions of said hard copy documents corresponding to said second data field being different from said portions of said hard copy documents corresponding to said first data field; and

storing information from said portions of said hard copy documents corresponding to said second data field into memory locations for said second data field.

24. A method as set forth in claim 23, further comprising the step of storing image information from said portions of said hard copy documents corresponding to said second data field into said memory locations for said second data field.

25. A method as set forth in claim 23, further comprising the step of storing textual information from aid portions of said hard copy documents corresponding to said second data field into said memory locations for said second data field.

26. A method as set forth in claim 23, further comprising the step of prompting identification of said portions of said hard copy documents corresponding to said second data field.

27. A method of processing data extracted from a diversity of types of hard copy documents, said method comprising the steps of:

(a) receiving output representing a diversity of types of hard copy documents from an automated digitizing unit and storing information from said diversity of types of hard copy documents into a memory as stored document information, said receiving step not preceded by scanning, via said automated digitizing unit, a separate document containing format requirements;

(b) recognizing and selecting portions of said stored document information corresponding to data fields required by an application unit;

(c) formatting said selected portions of said stored document information into a transmission format used by said application unit; and

(d) transmitting said formatted selected portions of said stored document information to said application unit.

28. A method as set forth in claim 27, further comprising the step of printing textual copies of said hard copy documents based on said stored document information.

29. A method as set forth in claim 27, wherein step (a) includes receiving output representing a diversity of types of hard copy documents from a scanner.

30. A method as set forth in claim 27, wherein step (a) includes receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a relative location on said hard copy documents containing data corresponding to a specific data field.

31. A method as set forth in claim 27, wherein step (a) includes receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a location on said hard copy documents containing data corresponding to a specific data field.

32. A method as set forth in claim 27, wherein step (a) includes the step of utilizing a template to associate portions of said hard copy documents with specific data fields.

33. A method as set forth in claim 27, further comprising detecting and correcting errors in said stored document information resulting from automated digitizing.

34. A method as set forth in claim 27, wherein step (a) includes storing textual information representing characters on said hard copy documents.

35. A method as set forth in claim 27, wherein step (a) includes storing digitized image information representing the actual appearance of said hard copy documents.

36. A method as set forth in claim 35, further comprising the step of printing copies of said hard copy documents based on said digitized image information.

37. An application program interface, comprising:
 an automated digitizing unit which extracts information from a diversity of types of hard copy documents and stores said information from said diversity of types of hard copy documents in a memory as stored document information;
 a processor recognizing and selecting portions of said stored document information corresponding to data fields required by an application unit, said processor operative without reference to information extracted, via said automated digitizing unit, from a separate document containing format requirements;
 a formatter formatting said selected portions of said stored document information into a transmission format used by said application unit; and
 an output unit transmitting said formatted selected portions of said stored document information to said application unit.

38. An interface as set forth in claim 37, wherein said automated digitizing unit includes a scanner.

39. An interface as set forth in claim 37, further comprising a search unit for searching for at least one character or symbol, located on said hard copy documents, which identifies a location on said hard copy documents containing data corresponding to a specific data field.

40. An interface as set forth in claim 37, further comprising:

41. An interface as set forth in claim 37, further comprising an error correcting unit detecting and correcting errors resulting from extracting by said automated digitizing unit.

42. An interface as set forth in claim 37, wherein said stored document information includes textual information representing characters on said hard copy documents.

43. An interface as set forth in claim 37, wherein said stored document information includes digitized image information representing an actual appearance of said hard copy documents.

44. An interface as set forth in claim 43, further comprising a printer which prints out copies of an actual appearance of said hard copy documents based on said digitized image information.

45. A method of processing data extracted from a diversity of types of hard copy documents, said method comprising the steps of:
 (a) receiving output representing a diversity of types of hard copy documents from an automated digitizing unit and storing information from said diversity of types of hard copy documents into a memory as stored document information, said receiving step not preceded by scanning, via said automated digitizing unit, of a separate document containing format requirements;
 (b) recognizing and selecting portions of said stored document information corresponding to data fields required by an application unit; and
 (c) transmitting said selected portions of said stored document information to said application unit.

46. A method as set forth in claim 45, wherein step (a) includes storing textual information representing characters on said hard copy documents.

47. A method as set forth in claim 45, further comprising detecting and correcting errors in said stored document information resulting from automated digitizing.

48. A method as set forth in claim 45, wherein step (a) includes the step of utilizing a template to associate portions of said hard copy documents with specific data fields.

49. A method as set forth in claim 45, wherein step (a) includes receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a location on said hard copy documents containing data corresponding to a specific data field.

50. A method as set forth in claim 45, wherein step (a) includes receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a relative location on said hard copy documents containing data corresponding to a specific data field.

51. A method as set forth in claim 45, further comprising the step of printing textual copies of said hard copy documents based on said stored document information.

52. A method as set forth in claim 45, wherein step (a) includes storing digitized image information representing an actual appearance of said hard copy documents.

53. A method as set forth in claim 52, further comprising the step of printing copies of said hard copy documents based on said digitized image information.

54. An application program interface, comprising:
an automated digitizing unit which extracts information from a diversity of types of hard copy documents and stores said information from said diversity of types of hard copy documents in a memory as stored document information;
a processor recognizing and selecting portions of said stored document information corresponding to data fields required by an application unit, said processor operative without reference to information extracted, via said automated digitizing unit, from a separate document containing format requirements; and
an output unit transmitting said selected portions of said stored document information to said application unit.

55. An application program interface as set forth in claim 54, wherein said automated digitizing unit includes a scanner.

56. An application program interface as set forth in claim 54, further comprising a search unit for searching for at least one character or symbol, located on said hard copy documents, which identifies a location on said hard copy documents containing data corresponding to a specific data field.

57. An application program interface as set forth in claim 54, further comprising a template definition unit for defining a template which associates locations on said hard copy documents with specific data fields.

58. An application program interface as set forth in claim 54, further comprising an error correcting unit detecting and correcting errors resulting from extracting by said automated digitizing unit.

59. An application program interface as set forth in claim 54, wherein said stored document information includes textual information representing characters on said hard copy documents.

60. An application program interface as set forth in claim 54, wherein said stored document information includes digitized image information representing an actual appearance of said hard copy documents.

61. An application program interface as set forth in claim 60, further comprising a printer which prints out copies of an actual appearance of said hard copy documents based on said digitized image information.

62. A method of processing data extracted from a diversity of types of hard copy documents, said method comprising the steps of:

(a) receiving output representing a diversity of types of hard copy documents from an automated digitizing unit and storing information from said diversity of types of hard copy documents into a memory as stored document information, said receiving step not preceded by scanning, via said automated digitizing unit, a separate document containing format requirements;

(b) recognizing and selecting portions of said stored document information corresponding to data fields required by an application unit; and

(c) formatting said selected portions of said stored document information into a transmission format used by said application unit.

63. A method as set forth in claim 62, further comprising the step of printing textual copies of said hard copy documents based on said stored document information.

64. A method as set forth in claim 62, wherein step (a) includes storing textual information representing characters on said hard copy documents.

65. A method as set forth in claim 62, wherein step (a) includes storing digitized image information representing an actual appearance of said hard copy documents.

66. A method as set forth in claim 62, further comprising detecting and correcting errors in said stored document information resulting from automated digitizing.

67. A method as set forth in claim 62, wherein step (a) includes the step of utilizing a template to associate portions of said hard copy documents with specific data fields.

68. A method as set forth in claim 62, wherein step (a) includes receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a location on said hard copy documents containing data corresponding to a specific data field.

69. A method as set forth in claim 62, wherein step (a) includes receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a relative location on said hard copy documents containing data corresponding to a specific data field.

70. An application program interface, comprising:
an automated digitizing unit which extracts information from a diversity of types of hard copy documents and stores said information from said diversity of types of hard copy documents in a memory as stored document information;
a processor recognizing and selecting portions of said stored document information corresponding to data fields required by an application unit, said processor operative without reference to extracted information scanned, via said automated digitizing unit, of a separate document containing format requirements; and
a formatter formatting said selected portions of said stored document information into a transmission format used by said application unit.

71. An application program interface as set forth in claim 70, wherein said stored document information includes textual information representing characters on said hard copy documents.

72. An application program interface as set forth in claim 70, further comprising an error correcting unit detecting and correcting errors resulting from extracting by said automated digitizing unit.

73. An application program interface as set forth in claim 70, further comprising a template definition unit for defining a template which associates locations on said hard copy documents with specific data fields.

74. An application program interface as set forth in claim 70, further comprising a search unit for searching for at least one character or symbol, located on said hard copy documents, which identifies a location on said hard copy documents containing data corresponding to a specific data field.

75. An application program interface as set forth in claim 70, wherein said automated digitizing unit includes a scanner.

76. An application program interface as set forth in claim 70, wherein said stored document information includes digitized image information representing an actual appearance of said hard copy documents.

77. An application program interface as set forth in claim 76, further comprising a printer which prints out copies of an actual appearance of said hard copy documents based on said digitized image information.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,258,855
DATED : November 2, 1993
INVENTOR(S) : Robert LECH et al.

Page 1 of 2

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 10, line 41, "read-in" should be --read in--.

Column 11, line 24, "maintains" should be --maintain--.

Column 16, line 27, Claim 1, "b" should be --by--, line 48, Claim 5, "sep" should be --step--.

Column 17, line 1, Claim 9, "potions" should be --portions--, line 6, Claim 9, "or" should be --for--, line 13, Claim 10, "potions" should be --portions--, line 38, Claim 14, "had" should be --hard--, line 42, Claim 14, "b" should be --by--.

Column 18, line 1, Claim 19, "Method" should be --method--, line 31, Claim 23, "potions" should be --portions--,

CERTIFICATE OF CORRECTION

PATENT NO. : 5,258,855
DATED : November 2, 1993
INVENTOR(S) : Robert LECH et al.

Page 2 of 2

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

line 44, Claim 25, "aid" should be --said--.

Signed and Sealed this
Twelfth Day of July, 1994

Attest:



BRUCE LEHMAN

Attesting Officer

Commissioner of Patents and Trademarks

United States Patent [19]

Lech et al.

[11] Patent Number: 5,369,508

[45] Date of Patent: * Nov. 29, 1994

[54] INFORMATION PROCESSING
METHODOLOGY

[75] Inventors: Robert Lech, Jackson; Mitchell A. Medina, Essex Fells; Catherine B. Elias, Plainsboro, all of N.J.

[73] Assignee: System X, L. P., New York, N.Y.

[*] Notice: The portion of the term of this patent subsequent to Nov. 2, 2010 has been disclaimed.

[21] Appl. No.: 143,135

[22] Filed: Oct. 29, 1993

Related U.S. Application Data

[63] Continuation of Ser. No. 672,865, Mar. 20, 1991, Pat. No. 5,258,855.

[51] Int. Cl.⁵ H04N 1/40

[52] U.S. Cl. 358/462; 358/448; 358/453; 358/449; 358/467; 382/61; 382/48

[58] Field of Search 358/462, 400, 401, 403, 358/447, 448, 449, 451, 452, 453, 460, 462, 463, 467, 470, 471, 474; 382/61, 48

[56] References Cited

U.S. PATENT DOCUMENTS

4,034,343	7/1977	Wilmer	340/146.3 MA
4,667,248	5/1987	Kanno	358/280
5,034,990	7/1991	Klees	382/22

5,095,445	3/1992	Sekiguchi	364/514
5,140,650	8/1992	Casey et al.	382/61
5,153,927	10/1992	Yamanari	382/61
5,258,855	11/1993	Lech et al.	358/462

Primary Examiner—Richard Hjerpe

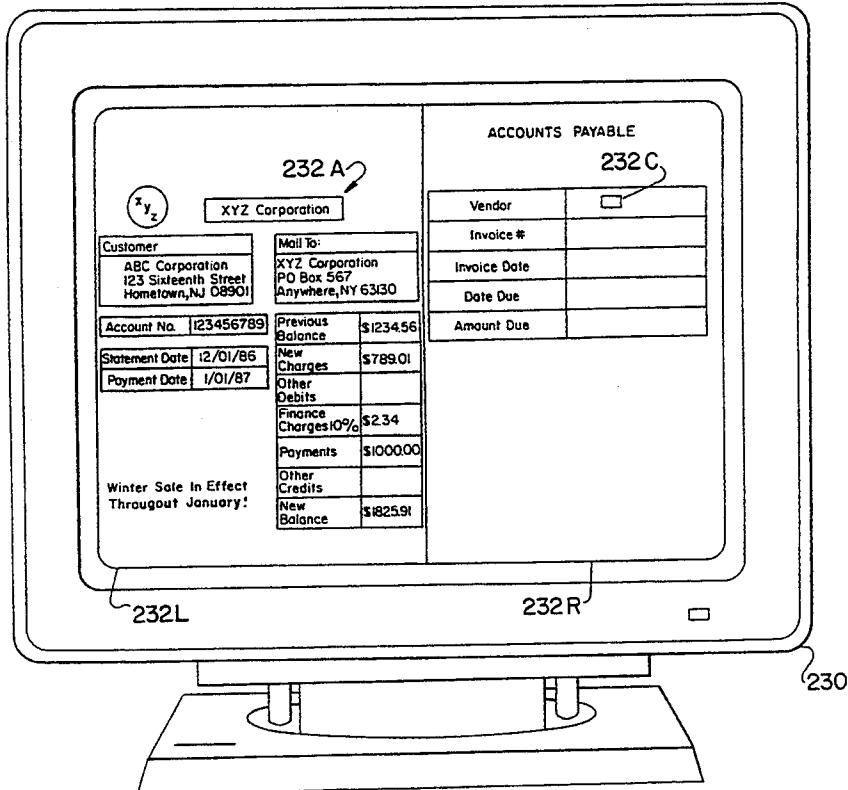
Assistant Examiner—Jerome Grant, II

Attorney, Agent, or Firm—Foley & Lardner

[57] ABSTRACT

An information processing methodology gives rise to an application program interface which includes an automated digitizing unit, such as a scanner, which inputs information from a diversity of hard copy documents and stores information from the hard copy documents into a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which designate portions of the stored document information required by a particular application program. The selected stored document information is then placed into the transmission format required by a particular application program in accordance with transmission format instructions. After the information has been transmission formatted, the information is transmitted to the application program. In one operational mode, the interface interactively prompts the user to identify, on a display, portions of the hard copy documents containing information used in application programs or for storage.

46 Claims, 15 Drawing Sheets



U.S. Patent

Nov. 29, 1994

Sheet 1 of 15

5,369,508

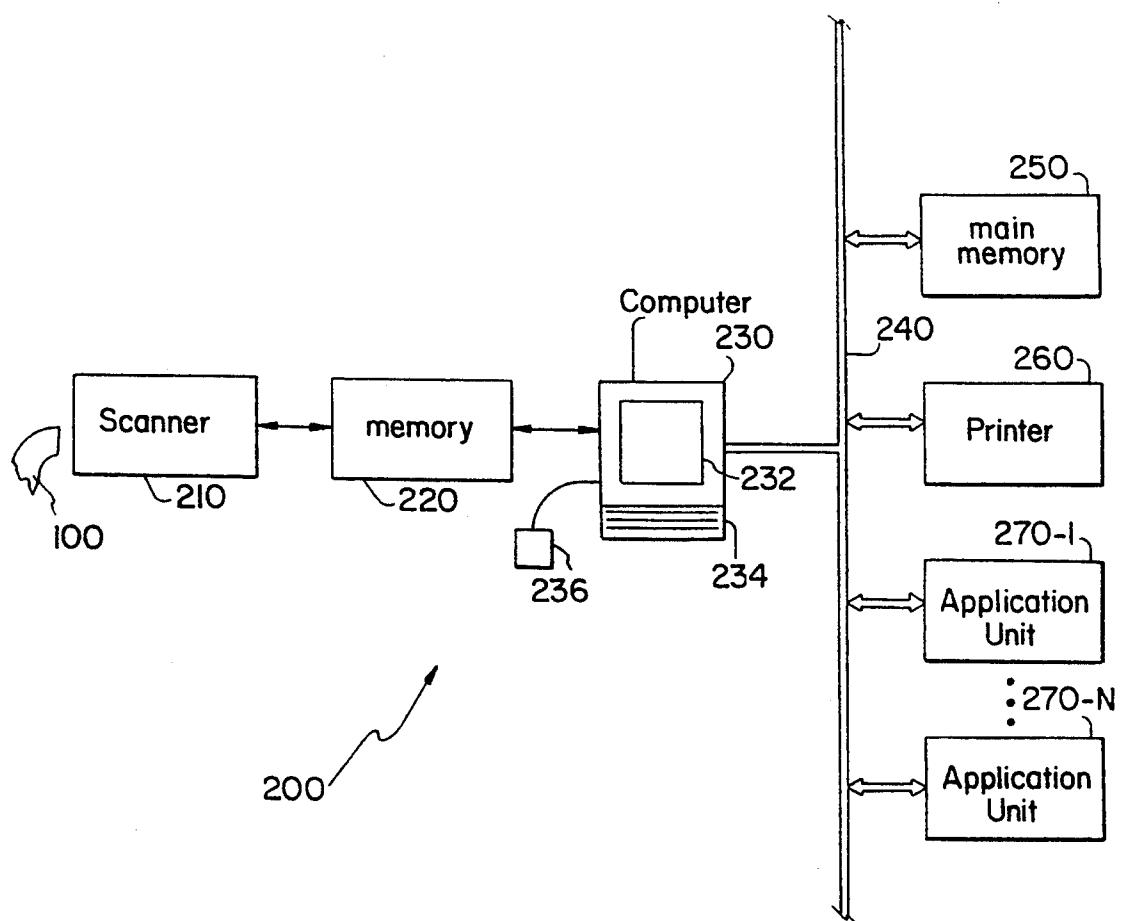


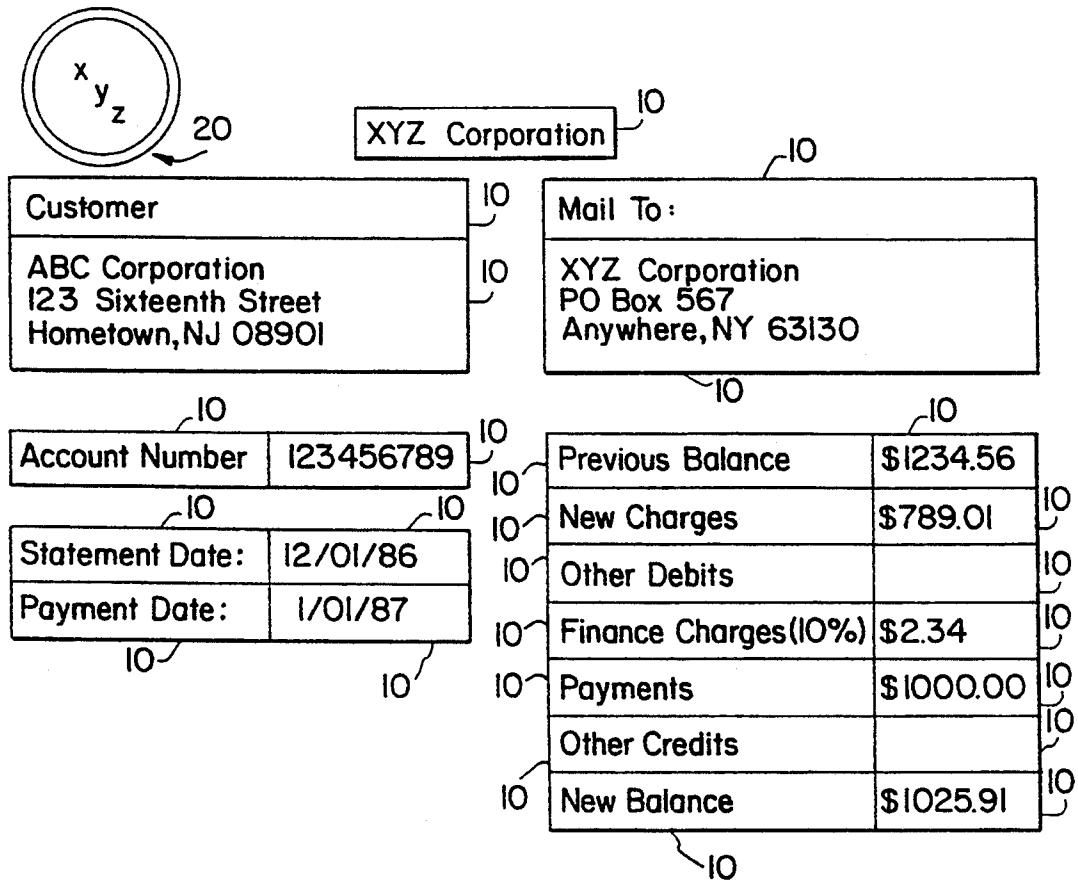
FIG. 1

U.S. Patent

Nov. 29, 1994

Sheet 2 of 15

5,369,508



Winter Sale In Effect Throughout January!

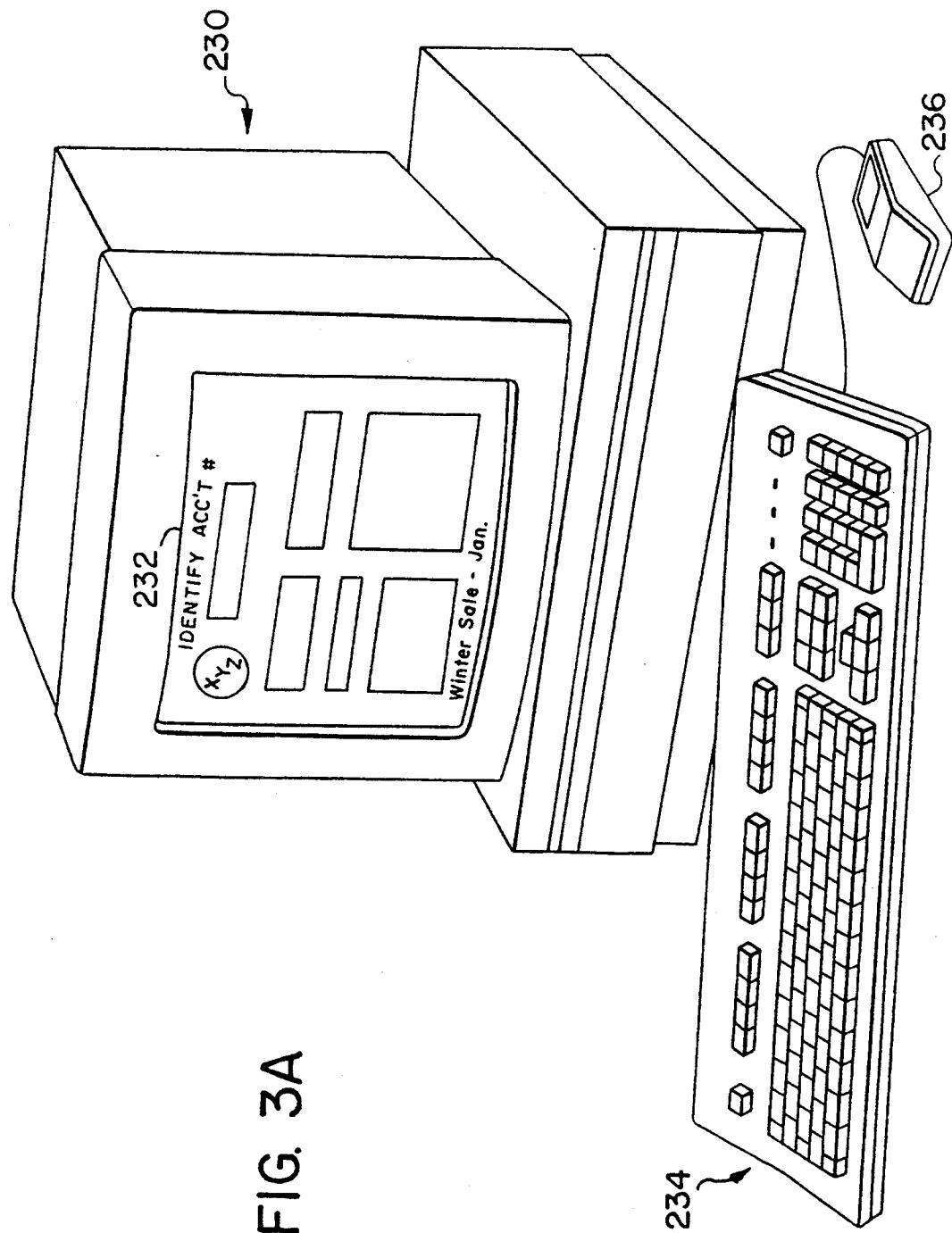
FIG. 2

U.S. Patent

Nov. 29, 1994

Sheet 3 of 15

5,369,508



U.S. Patent

Nov. 29, 1994

Sheet 4 of 15

5,369,508

232 A ↗

ACCOUNTS PAYABLE	
232 C	
Vendor	<input type="checkbox"/>
Invoice #	
Invoice Date	
Date Due	
Amount Due	
XYZ Corporation	<input type="checkbox"/>
Customer	
ABC Corporation	
123 Sixteenth Street	
HomeTown, NJ 08501	
Statement Date	12/01/86
Payment Date	1/01/87
Account No.	123456789
Previous Balance	\$1234.56
New Charges	\$789.01
Other Debits	
Finance Charges 0%	\$2.34
Payments	\$1000.00
Other Credits	
New Balance	\$1825.91
Winter Sale in Effect Throughout January!	

232 C ↗

232 R ↗

232 L ↗

230 ↗

FIG. 3B

U.S. Patent

Nov. 29, 1994

Sheet 5 of 15

5,369,508

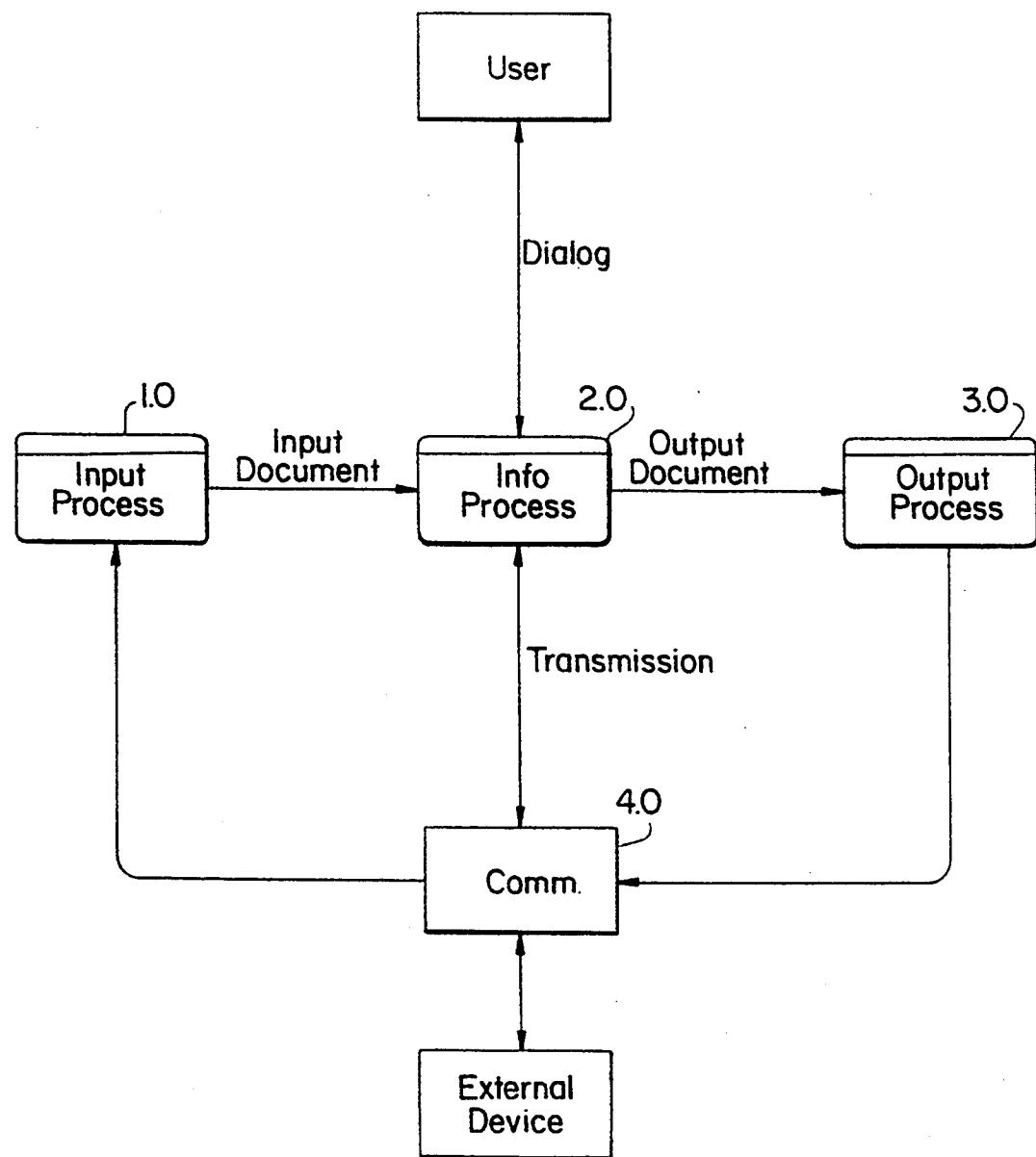


FIG. 4

U.S. Patent

Nov. 29, 1994

Sheet 6 of 15

5,369,508

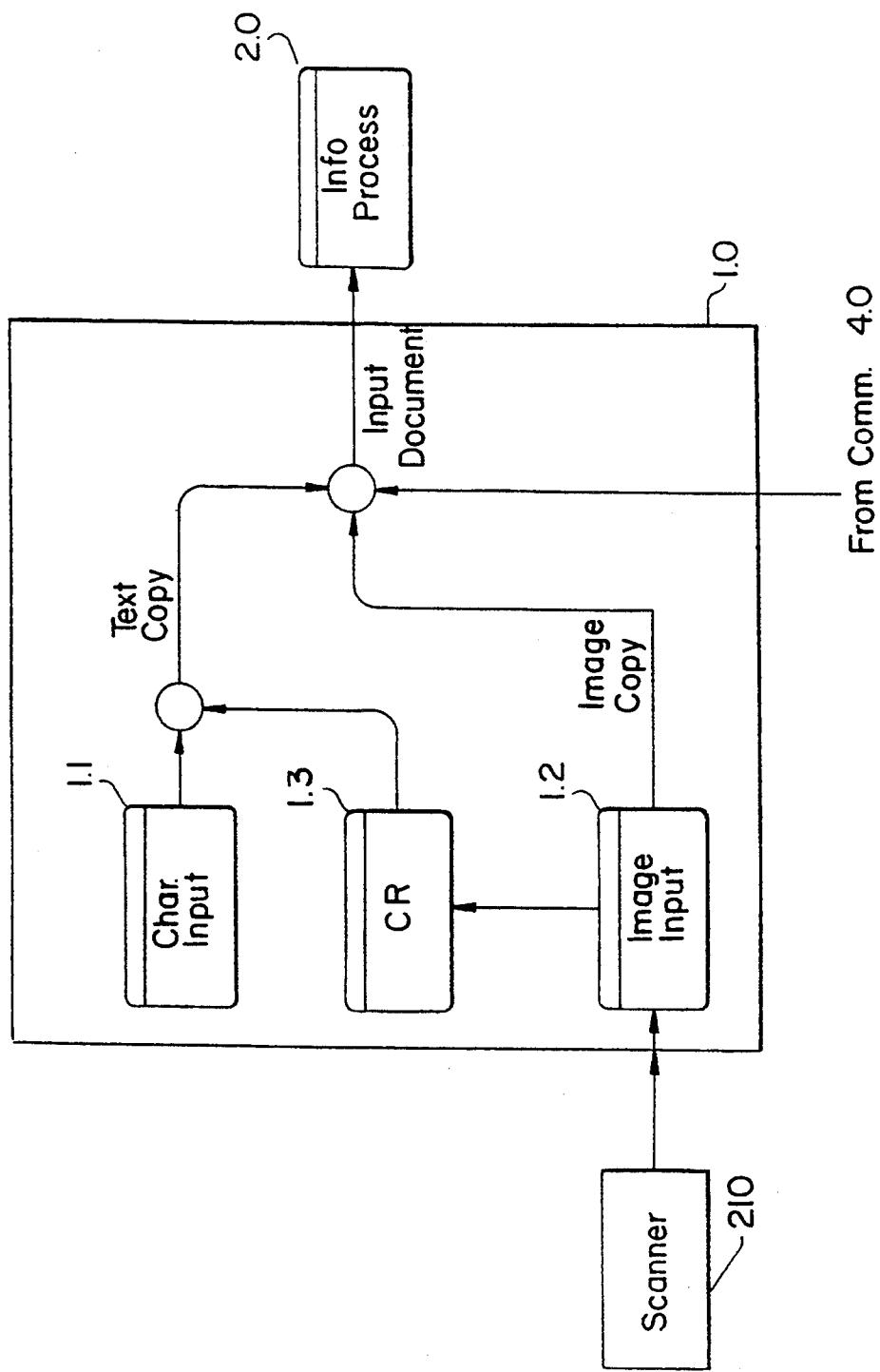


FIG. 5

U.S. Patent

Nov. 29, 1994

Sheet 7 of 15

5,369,508

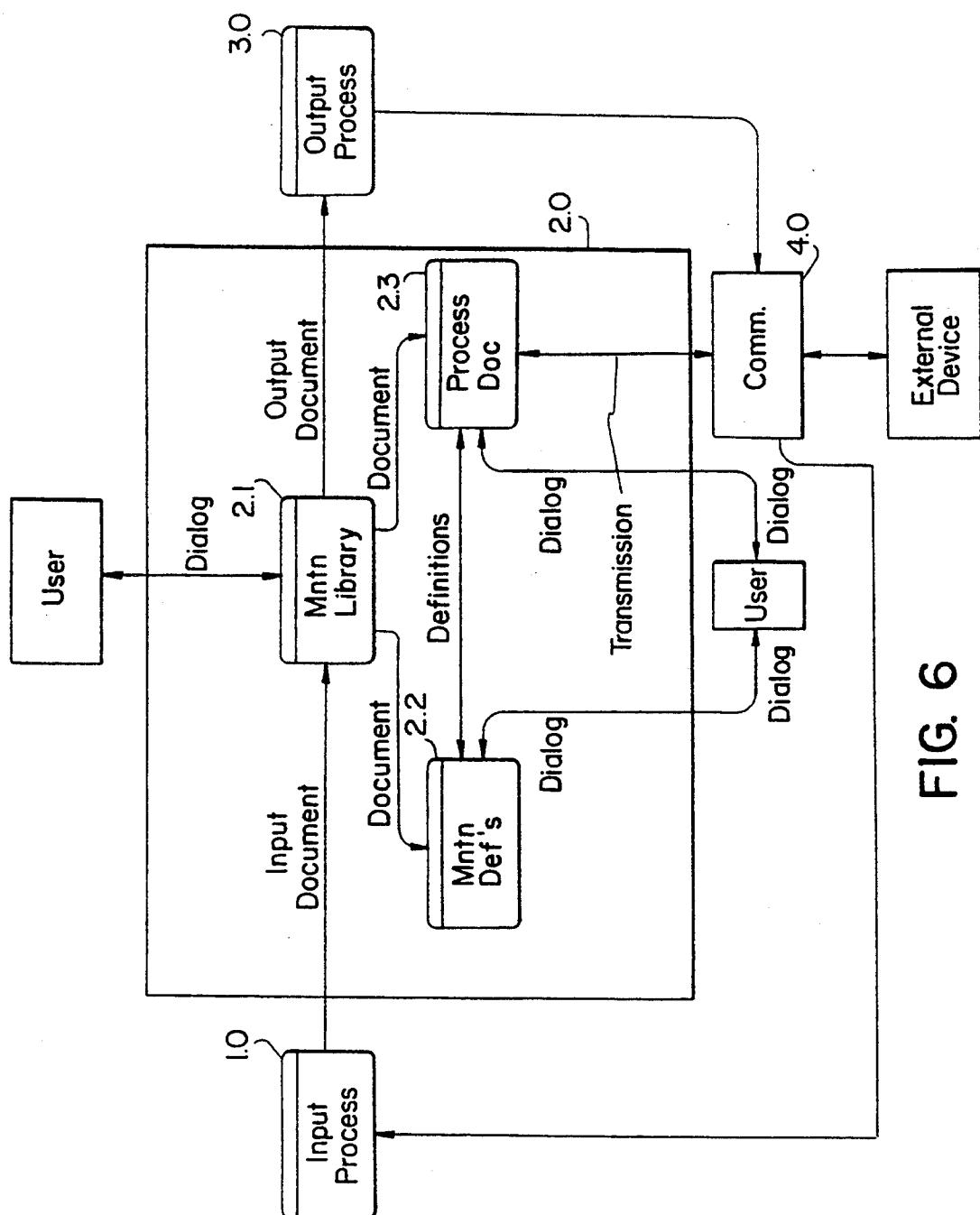


FIG. 6

U.S. Patent

Nov. 29, 1994

Sheet 8 of 15

5,369,508

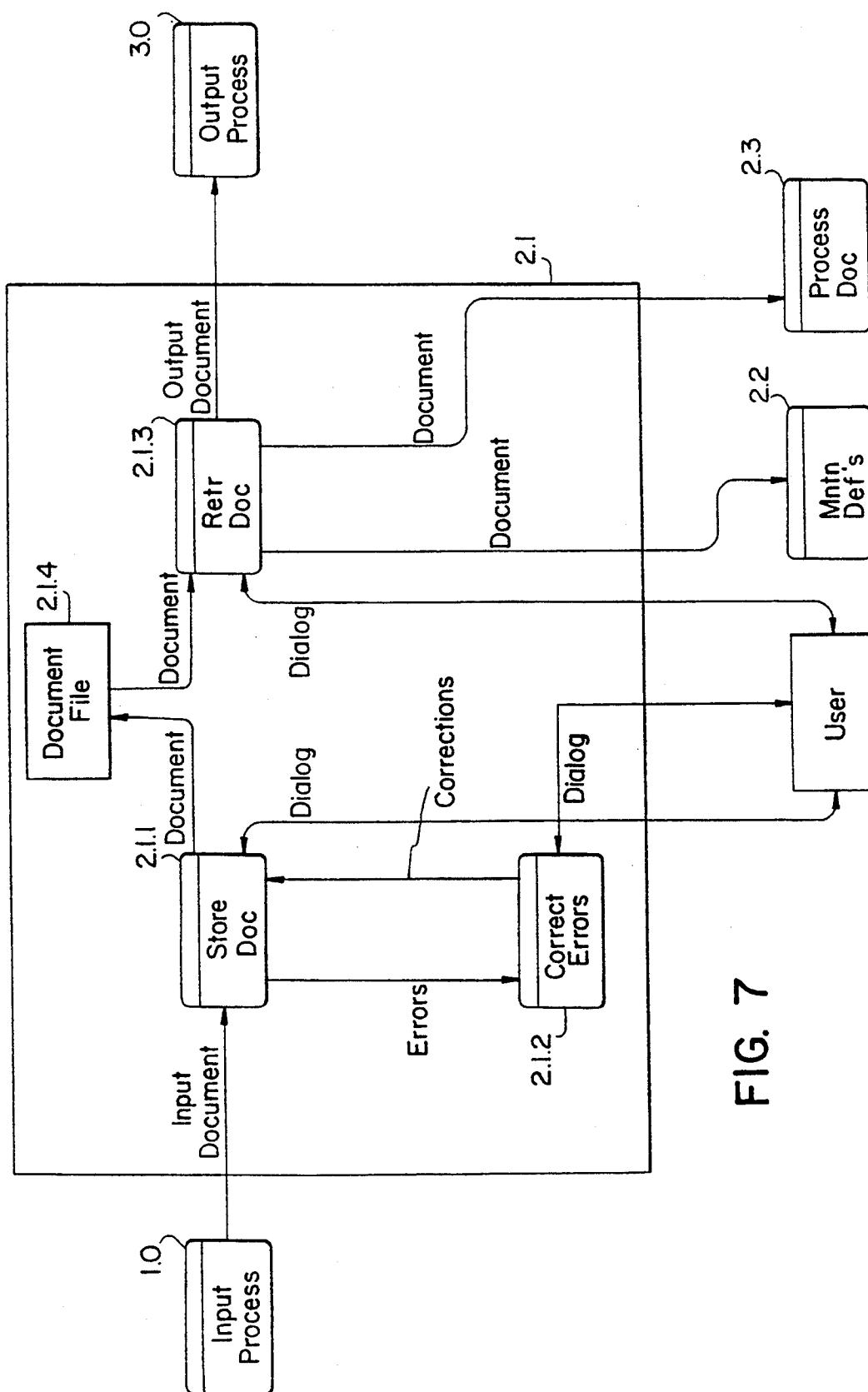


FIG. 7

U.S. Patent

Nov. 29, 1994

Sheet 9 of 15

5,369,508

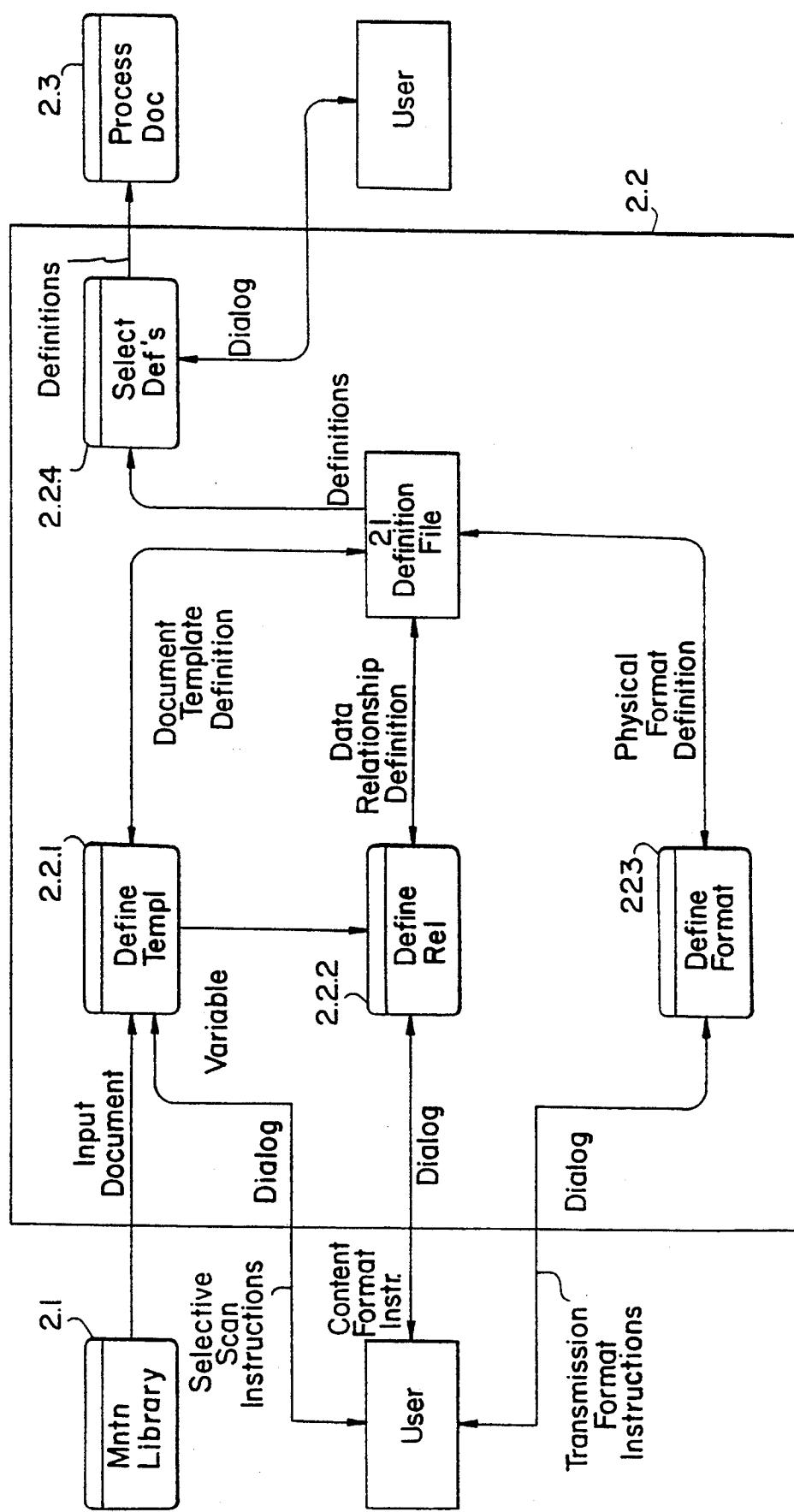


FIG. 8

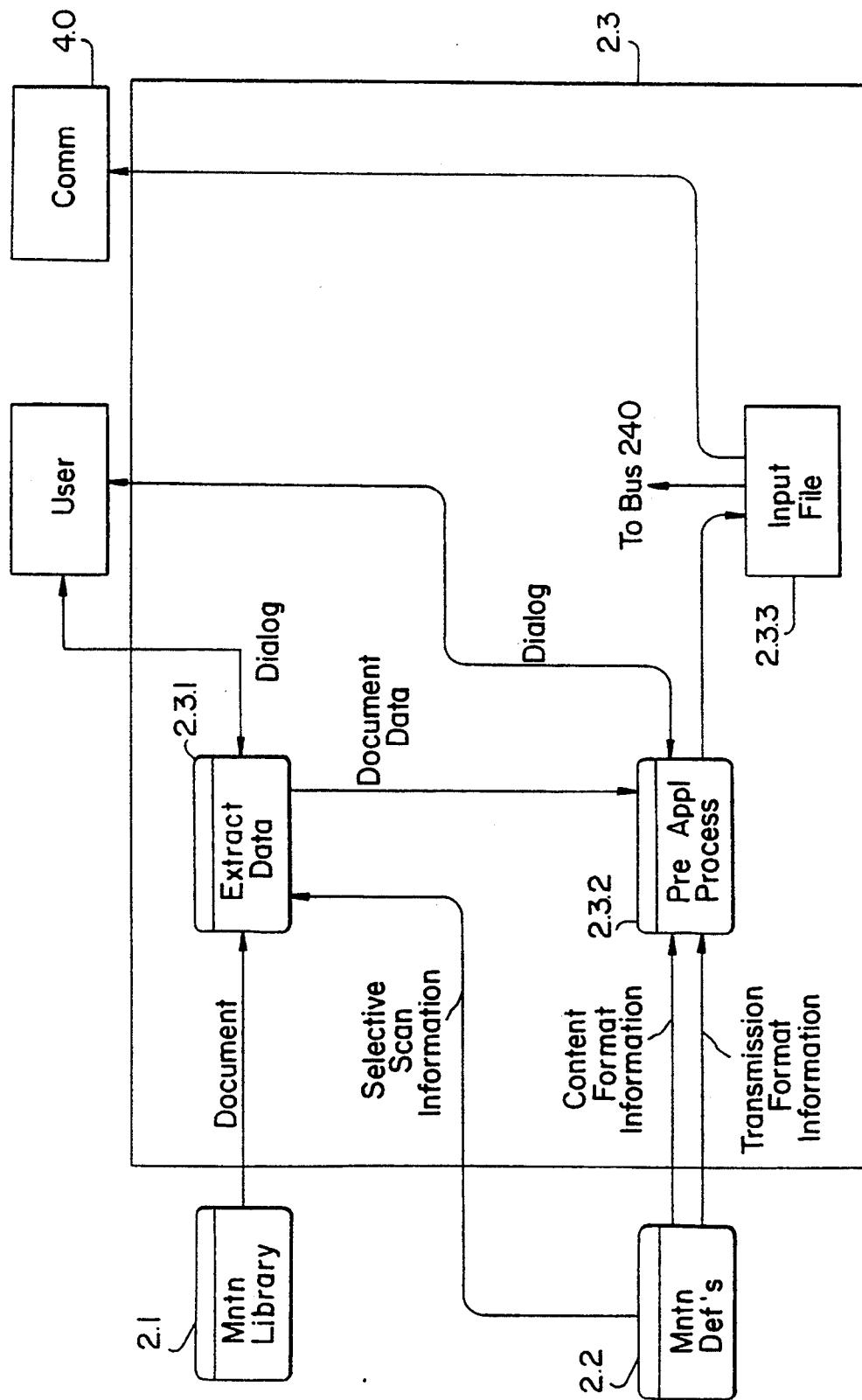


FIG. 9

U.S. Patent

Nov. 29, 1994

Sheet 11 of 15

5,369,508

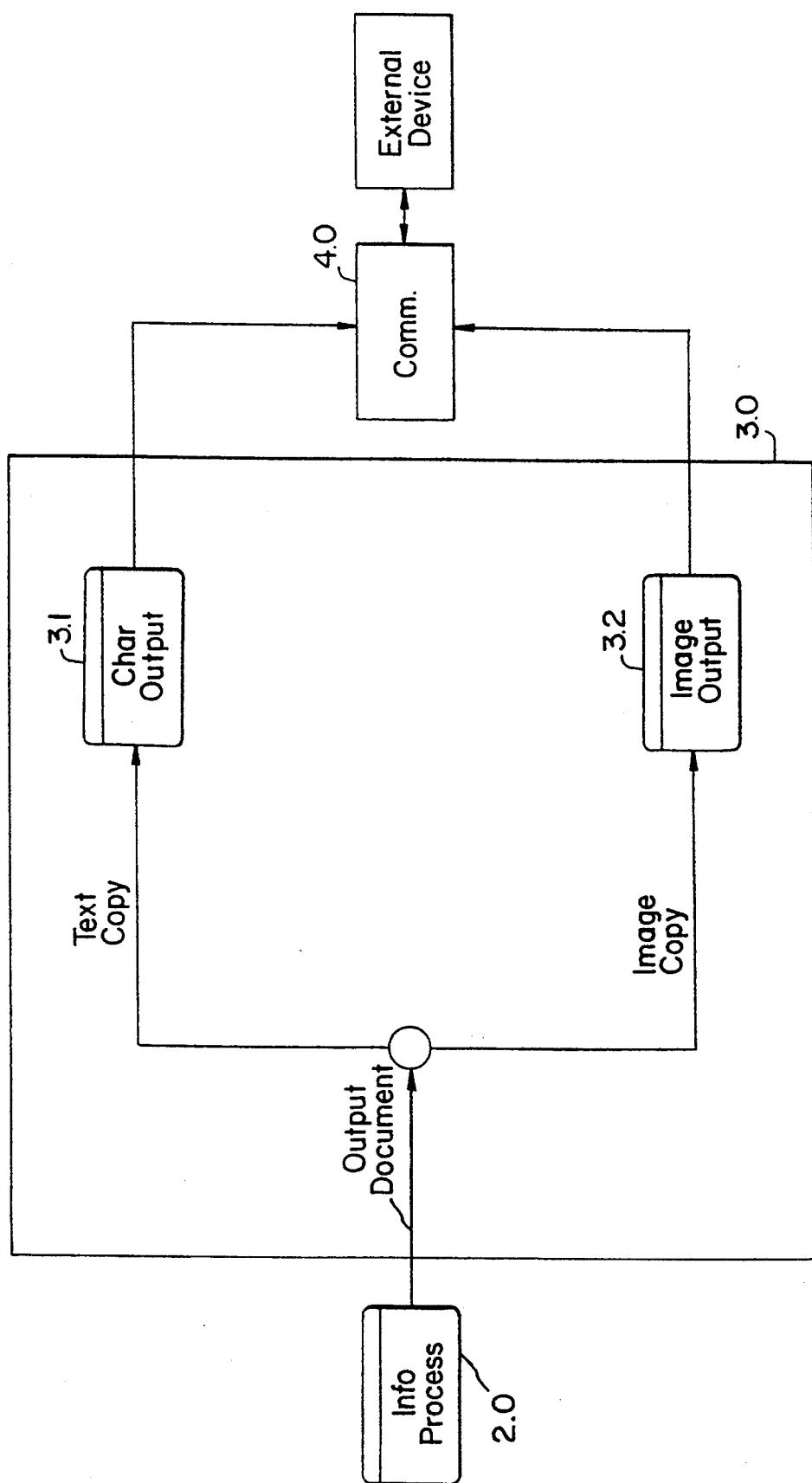


FIG. 10

U.S. Patent

Nov. 29, 1994

Sheet 12 of 15

5,369,508

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Heading 2	2
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
Other Credits	
New Balance	\$1025.91

FIG. 11

U.S. Patent

Nov. 29, 1994

Sheet 13 of 15

5,369,508

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
New Balance	\$1025.91

FIG. 12A

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

FIG. 12B

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

FIG. 12C

U.S. Patent

Nov. 29, 1994

Sheet 14 of 15

5,369,508

```
>1>1>"XYZ Corporation"  
>2>2>25>+123456789>  
>3>2>1>D12/01/86>  
>4>2>11>D12/15/86>  
>5>2>21>D01/01/87>  
>6>10>25>$1234.56  
>7>11>25>$789.01>  
>8>13>25>$2.34>  
>9>14>25>$1000.00>  
>10>16>25>$1025.91>
```

FIG. 13A

```
>1>1>1>$1234.56>  
>2>2>1>"XYZ Corporation"  
>3>3>1>"PO Box 567"  
>4>4>1>"Anywhere, NY 63130"
```

FIG. 13B

```
/1/1/1/$1234.56//  
/2/2/1/*XYZ Corporation*  
/3/3/1/*PO Box 567*  
/4/4/1/*Anywhere, NY 63130*
```

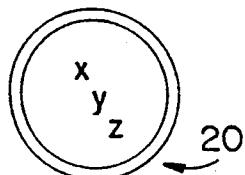
FIG. 13C

U.S. Patent

Nov. 29, 1994

Sheet 15 of 15

5,369,508



XYZ Corporation

Customer

ABC Corporation
123 Sixteenth Street
Hometown, NJ 08901

Mail To:

▲XYZ Corporation
PO Box 567
Anywhere NY, 63130▲

Account Number 123456789

Previous Balance \$1234.56

Statement Date 12/01/86

New Charges \$789.01

Payment Date 1/01/87

Other Debits

Finance Charges (10%) \$2.34

Payments \$1000.00

Other Credits

New Balance \$1025.91

Winter Sale in Effect Throughout January

30

FIG. 14

INFORMATION PROCESSING METHODOLOGY

This is a continuation of application Ser. No. 07/672,865, filed Mar. 20, 1991 now U.S. Pat. No. 5,258,855.

BACKGROUND OF THE INVENTION

The invention is directed to a system for efficiently processing information originating from hard copy documents. More specifically, the invention is directed to a hard copy document to application program interface which minimizes the need to manually process hard copy documents.

In the past, information contained on hard copy documents was manually entered into a computer via the input controller of a particular computer. The original document was then filed away for future reference. Automatic input of data was limited to the input of Magnetic Ink Character Recognition (MICR) data and to Optical Character Recognition (OCR) data. This fixed-position data was forwarded directly to a dedicated computer application specifically designed to accommodate the input format. In more recent years, typewritten text has been mechanically inputted into a computer via a text file. Examples of this latter type of system are word processors and photo-typesetters

These conventional systems have limitations which decrease the efficiency of processing information from a hard copy document. For example, the systems discussed above are limited in their application to MICR, OCR, or typewritten data. Parsing and processing data is limited to the particular requirements of the particular computer application which requires the input data. In addition, in these conventional systems, the actual hard copy document must be retained for future reference at great expense.

In a sophisticated computer network, different users may require different portions of the information contained on a hard copy document. For example, if the hard copy document is an invoice returned with payment of a bill, the accounting department may need all of the monetary information contained on the bill while the mailroom may need only customer address information, to update a customer's address. Therefore, there is a need for a system in which specific information from a hard copy document can be selectively distributed to various users.

Another problem with conventional systems is that users, even within the same company, may require that the information extracted from a hard copy document be transmitted to a particular application program in a specific transmission format. For example, one department in a company may use a particular application program which must receive information using a particular character as a delimiter and other departments may require the information in a different format using different delimiters.

Another problem, particularly for small businesses, is that current systems can not efficiently accommodate the inputting of information from a diversity of hard copy documents. A large business which receives many forms in the same format can afford a system which inputs a high volume of information in that format into memory. For example, it is cost-effective for a bank which processes hundreds of thousands of checks a month to buy a dedicated machine which can read information off of checks having a rigidly defined, or

fixed, format. However, as the diversity of forms received by a business increases relative to the number of forms that must be processed, it becomes less cost-effective to design a dedicated machine for processing each type of form format. This problem is particularly significant in small businesses which may, for example, receive fifty invoices a month, all in different, non-fixed, formats. It is frequently not cost-effective for a small business to design dedicated systems for inputting information in each of these various formats. This leaves a small business with no other practical alternative than to manually input the information off of each invoice each month.

SUMMARY OF THE INVENTION

It is an object of the invention, therefore, to provide an application program interface which allows a user to select specific portions of information extracted from a diversity of hard copy documents and allows the user to direct portions of this information to several different users in accordance with the needs of the particular user.

It is also an object of the invention to provide a cost-effective system for inputting hard copy documents which can accommodate hard copy documents in a diversity of formats.

It is another object of the invention to provide an application program interface which allows a user to put information, which is to be transmitted, into a particular transmission format, based upon the needs of the receiver of the information.

It is a further object of the invention to provide an application program interface which will allow the extraction, selection, formatting, routing, and storage of information from a hard copy document in a comprehensive manner such that the hard copy document itself need not be retained.

It is another object of the invention to provide a system which reduces the amount of manual labor required to process information originating from a hard copy document.

A further object of the invention is to reduce the time required to process information originating from a hard copy document so that a higher volume of transactions involving hard copy documents can be processed.

The invention provides an application program interface which inputs a diversity of hard copy documents using an automated digitizing unit and which stores information from the hard copy documents in a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which define portions of the stored document information required by a particular application unit. Selected stored document information is then formatted into the transmission format used by the particular application program based on transmission format instructions. The transmission formatted selected stored document information is then transmitted to the particular application program. The hard copy documents may contain textual information or image information or both.

The interface operates in three different modes.

In a first mode, the interface extracts all of the information from hard copy documents and stores this information in memory. Parsing of various portions of the extracted information is performed in accordance with content instructions.

In a second mode, the user operates interactively with the interface by use of a display and an input device, such as a mouse. In this second mode, a hard copy document is inputted and displayed on the display. The interface then prompts the user to identify the location of various information. For example, the interface can ask the user to identify the location of address information on the hard copy document. In response, the user positions the mouse to identify address information using a cursor. The identified information is then stored as address information in memory. Subsequently, the interface again prompts the user to identify other pieces of information, which are then stored in the appropriate locations in memory. This process proceeds until all of the information which is desired to be extracted off of the hard copy document is stored in memory. 15

In a third mode of operation, selected portions of information are extracted off of hard copy documents in accordance with predetermined location information which has been specified by the user. For example, the user can define a template which specifies the location of information on hard copy documents. Templates can be formed in conjunction with second mode operation. Alternatively, the user can instruct the interface to search hard copy documents for a particular character or symbol, located on the hard copy documents. The information desired to be extracted off of the hard copy documents is specified relative to the location of this character or symbol. 25

The interface can also prompt or receive from an applications program or another information processing system, required information, content instructions, and format instructions. 30

Other objects, features, and advantages of the invention will be apparent from the following detailed description of the invention. 35

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be described in further detail below with reference to the accompanying drawings, in which:

FIG. 1 illustrates hardware for implementing a preferred embodiment of the instant invention; 40

FIG. 2 illustrates an example of a hard copy document containing information to be processed by the instant invention; 45

FIGS. 3A and 3B are enlarged views of the computer of FIG. 1 used to explain how the invention interactively prompts a user to identify information; 50

FIG. 4 is an overall data flow diagram for the FIG. 1 preferred embodiment;

FIG. 5 is a detailed input data flow diagram for the FIG. 1 preferred embodiment;

FIG. 6 is a detailed information processing data flow diagram for the FIG. 1 preferred embodiment; 55

FIG. 7 is a more detailed information processing data flow diagram for the maintain library module of FIG. 6;

FIG. 8 is a more detailed information processing data flow diagram for the maintain definitions module of FIG. 6; 60

FIG. 9 is a more detailed information processing data flow diagram for the process document module of FIG. 6;

FIG. 10 is a detailed output data flow diagram for the FIG. 1 preferred embodiment; 65

FIG. 11 lists data corresponding to the hard copy document of FIG. 2;

FIGS. 12A, 12B, and 12C illustrate examples of data which can be selected from the extracted data of FIG. 11 in accordance with content instructions;

FIGS. 13A, 13B, and 13C illustrate examples of the data of FIGS. 12A, 12B, and 12C formatted in accordance with various transmission format instructions to form input files; and

FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention. 10

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Hardware

The invention provides an interface between information originating from a hard copy document and a computer application unit which uses the information. The computer application unit can be a particular computer application program or a device which is controlled in accordance with instructions or information from the hard copy document. The invention also allows storing a copy of the hard copy document in a memory and retrieving the copy of the hard copy document. By providing a comprehensive and integrated system which can accommodate almost all of the possible uses of information contained on a hard copy document, the instant invention allows for a paperless office. 20

The invention includes hardware and software necessary to extract, retrieve, and process information from the hard copy document. A copy of the actual image of the hard copy document is stored in memory. Textual information extracted from the hard copy document is also stored in memory. Textual information is information, such as alphanumeric characters, which is recognized on the hard copy document and which is stored in a form which corresponds to the particular recognized character. For example, the extracted characters can be stored in the ASCII format in an electronic memory. 30

The user can have all of the information extracted from the hard copy document and stored in memory. Alternatively, the interface can interactively prompt the user to identify specific pieces of information for storage. The interface can also extract specific pieces of information using a predefined template. The interface can also prompt or receive from another information processing system or an applications program desired information, content instructions, and format instructions. 40

The instant invention also provides for parsing information extracted from the hard copy document and for directing this parsed information to specific users or application programs as an input file. 50

The invention also permits the user to define the transmission format of the input file for a particular computer application unit.

FIG. 1 illustrates hardware for implementing a preferred embodiment of a hard copy document to application program interface according to the instant invention. The interface 200 processes information extracted off of hard copy document 100 and provides information to application units 270 in a form required by each particular application unit. The interface extracts information off of a hard copy document 100 utilizing a scanner 210. The scanner 210 can be any type of scanner which extracts information off of hard copy documents, for example, an Optical Reader. 60

The scanned information is stored in a scanner memory 220 or in main memory 250, as will be described in greater detail below. If main memory 250 or another memory is available to store the scanned information, then scanner memory 220 can be omitted.

The information from scanner memory 220 or main memory 250 is transmitted to computer 230. In the preferred embodiment, computer 230 includes a display 232, a keyboard 234, and a mouse 236. The display 232 displays an image of the hard copy document itself and/or information necessary to process the information extracted off of the hard copy document.

The computer 230 is used to select portions of the stored document information contained in memory in accordance with content instructions which define portions of the stored document information required by an application unit. These content instructions may be provided by the application program. Alternatively, the content instructions can be inputted via an input device such as a keyboard, a touch screen, a mouse, a notepad or a voice recognition device, or the like.

The computer 230 is also used to format selected stored document information into the transmission format used by an application unit based on transmission format instructions. The transmission format instructions may be provided by the application program. Alternatively, the transmission format instructions can be inputted via a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

Thus, the computer 230 is used to generate an input file for a particular application unit. The computer 230 is connected to scanner memory 220, main, or permanent, memory 250, a printer 260, and application units 270, via a bus 240. Although FIG. 1 illustrates use of a bus to connect components together, it is understood that any routing or connecting link, implemented in hardware or software or both, can be employed instead of, or in addition to, a bus. Instructions to or in the computer 230 control the main memory 250, the printer 260, the application units 270, and the bus 240. Instructions to or in computer 230 can also control exchanges of information with scanner memory 220.

When the computer 230 generates an input file for a particular document, the computer 230 can send this input file directly to an application unit or can store this input file in the main memory 250 until required by an application unit. The main memory 250 may also optionally store a copy of the image information for the hard copy document and the textual information for the hard copy document. Thus, the image information and textual information from the hard copy document can be retrieved and printed out on printer 260. In addition, image and textual information stored in scanner memory 220 or in main memory 250 can be used to form additional input files at the time of input or at a later time, based on content instructions and transmission format instructions. Thus, the invention can, at the discretion of the user, eliminate the need to retain copies of hard copy documents, permitting a paperless office.

The application units 270 include particular application programs and devices which are controlled in accordance with information contained on hard copy document 100.

FIG. 2 illustrates an example of a hard copy document 100 which contains information to be processed by the instant invention. The document illustrated in FIG. 2 is a bill from XYZ Corporation to customer ABC Corporation. FIG. 2 is only an example of a type

of document that can be processed by the instant invention.

In a first operational mode, the scanner 210 stores all of the information extracted off of hard copy document 100 in the scanner memory 220 or, alternatively, in main memory 250. The extracted information is stored in two forms. The actual image of the hard copy document 100 is stored as image information in the scanner memory 220. In addition, the scanner memory 220 stores textual information recognized on the hard copy document 100 by, for example, employing standard character recognition software. In the preferred embodiment, the textual information is stored in ASCII format. The scanner memory 220 can be, for example, an electronic, magnetic, or optical memory.

FIG. 3A illustrates an enlarged view of the computer 230 of FIG. 1. This view will be used to describe a second mode of operation. In this second mode of operation, the hard copy document 100 is scanned and a copy of the document 100 is displayed on display 232 of computer 230, based on the contents of information temporarily stored in scanner memory 220. After the document is displayed on display 232, the computer 230 interactively prompts the user to identify the location of specific pieces of information on the hard copy document. In the FIG. 3A illustration, this prompt message is indicated as the message beginning with the arrow.

For example, the prompt message can ask the user to identify the location of account number information on the hard copy document. The user then uses an input device, such as keyboard 234 or mouse 236 or a touch screen, notepad, voice recognition device, or other input device to position a cursor on the display to identify the location of the information requested by the prompt message. For example, the cursor could be used to define a block (which could be highlighted) containing the requested information, followed by a mouse "enter" click. In this example, the user would move the mouse to identify the location of the account number information contained on the hard copy document 100. The computer 230 then stores the information which has been identified by the user as account number information in the appropriate address or subfile or as the appropriate variable or parameter in memory. The computer then prompts the user to identify the location of other information on the hard copy document, such as, statement date information. The process proceeds until all of the desired information has been stored into the appropriate locations in memory.

FIG. 3B illustrates a variation of the second mode for interactively prompting the user for information. In FIG. 3B, the display is split into two portions. A left-hand portion 232L displays the image of the hard copy document and a right-hand portion 232R displays the required application program information. For example, in FIG. 3B, portion 232R displays a spreadsheet used by an application program. While observing the split display, the user can input instructions to associate specific pieces of information on the hard copy document (for example, the vendor name indicated by the mouse arrow 232A) with particular subfiles in memory (for example, the vendor field next to which the cursor 232C appears), using a mouse or other input device(s) or both. The split display also allows the user to generate content format instructions while observing the information required for a particular application program on the right-hand portion.

These second modes of operation are efficient for small businesses which receive a small number of a wide variety of invoices, since the user does not necessarily have to store all of the information that appears on the hard copy document. A further advantage is that data input is quicker, easier, and more accurate than with previous keyboard methodology. In addition, by specifying the location on the hard copy document of information, the user may optionally create a template, to be described in further detail below, for each different type of invoice. This template is stored for future use when another hard copy document in the same format is received.

More specifically, instructions from computer 230 can direct the scanner 210 and scanner memory 220, and/or main memory 250, to scan and/or store only specific portions of hard copy document 100. After the interactive prompts required to obtain information for a desired application program, the unused information stored in scanner memory 220 or 250 can be erased. Further, scanning of a second identical document can be limited to only those portions of the document which contain needed information.

More specifically, in FIG. 2, the lines 10 drawn around certain portions of the document represent the areas which the user has previously identified as the portions of a document to be extracted by the scanner 210 and stored in scanner memory 220 and/or main memory 250. Since the logo 20 and the message 30 have not been identified as an area to be scanned and stored, these areas are not scanned and stored in subsequent documents. Since the user has previously associated each of the areas 10 with a specific subfile of information, e.g., the account number, the scanned information is stored in memory locations corresponding to that subfile.

Data Processing

FIGS. 4-10 illustrate the flow of data in the FIG. 1 preferred embodiment. FIG. 4 illustrates the overall data flow for the FIG. 1 preferred embodiment. The preferred embodiment includes an input process module 1.0, an information processing module 2.0, and an output processing module 3.0. The information processing module 2.0 is equipped to receive instructions from and transmit information to a user. The information processing module 2.0 can also transmit to and receive information from a remote external device through communication interface 4.0. Input process module 1.0 and output processing module 3.0 can also access communication interface 4.0. A module is implemented in hardware, software, or a combination of hardware and software. The specific implementation for a particular business application depends upon a variety of factors, for example, the relative costs of hardware and software implemented systems, the frequency with which a user will want to expand or modify the system, and the like.

FIG. 5 is a more detailed diagram of the input process module 1.0 of FIG. 4. The input process module 1.0 includes a character input module 1.1, an image input module 1.2, and, in the preferred embodiment, a character recognition device 1.3. The character input module inputs textual information, such as alphanumeric characters, from

an input device such as keyboard 234. The image input module 1.2 inputs image information, for example, a digitized image of the actual appearance of hard copy document 100. Textual information can include textual

input from an input device such as keyboard 234 and textual information extracted from the document by character recognition device 1.3. Both types of information comprise an input document which is transmitted to information processing module 2.0. In the FIG. 1 preferred embodiment, the processing performed by input process module 1.0 occurs in scanner memory 220, computer 230, and main memory 250.

FIG. 6 illustrates information processing data flow for the FIG. 1 preferred embodiment, that is, FIG. 6 illustrates data flow in the information processing module 2.0.

The information processing module 2.0 includes a maintain library module 2.1, to be described in further detail below in conjunction with FIG. 7, a maintain definitions module 2.2, to be described in further detail below in conjunction with FIG. 8, and a process document module 2.3 to be described in further detail below in conjunction with FIG. 9.

The information processing module 2.0 is the module which coordinates and drives the entire system. In the preferred embodiment, the information processing module 2.0 is implemented primarily by computer 230.

FIG. 7 illustrates information processing data flow in the maintain library module 2.1. The maintain library module 2.1 maintains a library of image information, for example, a digitized image representing the actual appearance of the hard copy document, and textual information of the hard copy documents for reference during processing. This library can be incorporated within scanner memory 220, main memory 250, or another independent memory, for example, a RAM disk. The maintain library module 2.1 includes a store document module 2.1.1, a correct errors module 2.1.2, a retrieve document module 2.1.3, and a document file 2.1.4. These modules operate collectively to store, retrieve, and correct document information.

The store document module 2.1.1, prior to routing the document to the document file 2.1.4, may provide information on recognition errors which may have occurred while inputting the document. For example, the store document module 2.1.1 identifies that a character contained on hard copy document 100 was not recognized. The store document module 2.1.1 also optionally causes a copy of the document and its parsing to be displayed on the display 232 for confirmation by the user. The user may utilize this opportunity to identify any errors in the displayed document and, in conjunction with the correct errors module 2.1.2, to revise the document's parsing, if necessary, prior to storage of the document in memory. The module 2.1.1 also provides a facility for the user to name a particular hard copy document for cataloging, storage, and retrieval purposes. After the document is named, the store document module 2.1.1 stores copies of the document in the document file 2.1.4.

The correct errors module 2.1.2 processes instructions from the user to correct errors identified by the store document module 2.1.1 and errors that have been spotted by the user during the confirmation process.

The retrieve document module 2.1.3 permits the user to retrieve a copy of a document previously stored in the document file 2.1.4. As described above, long-term storage is provided by main memory 250, if necessary.

FIG. 8 illustrates a more detailed information processing data flow diagram for the maintain definitions module 2.2 of FIG. 6. The maintain definitions module 2.2 allows the user to define system and document pa-

rameters and maintains the definitions of these system and document parameters. The maintain definitions module 2.2 includes a define template module 2.2.1 which allows the user to specify the location of information on the document. This information provided by the user defines a template which is used to extract information off the document and to associate the extracted information with a particular variable or subfile. These templates are illustrated by boxes 10 in the FIG. 2 example of a hard copy document. The maintain definitions module 2.2 can also access templates previously defined by the user and stored in main memory 250. Templates can also be provided as part of software packages developed by program developers.

The maintain definitions module 2.2 also includes a define relationships module 2.2.2. The define relationships module 2.2.2 allows the user to define data relationships, or logical relationships, between pieces of information extracted from the hard copy document. These pieces of information are then used to generate an input file for a selected computer application unit. The user defines these relationships by content instructions. Alternatively, content instructions to define relationships can be provided by application software. If the user provides these content instructions, the content instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. Examples of content instructions, data, and logical relationships will be described in further detail in conjunction with FIGS. 11 and 12A, 12B, and 12C.

The maintain definitions module 2.2 also includes a define format module 2.2.3. The define format module 2.2.3 allows the user to define transmission formats for an input file which is then transmitted to a selected computer application unit. Selection of the transmission format of the input file is accomplished by the user through use of transmission format instructions. Alternatively, the applications software itself can generate its own transmission format instructions. When the user must specify transmission format instructions, the transmission format instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. A further description of various transmission formats will be provided below in conjunction with FIGS. 12A, 12B, 12C, 13A, 13B, and 13C.

A select definitions module 2.2.4 is also included in the maintain definitions module 2.2. The select definitions module 2.2.4 allows the user to store and select a set of definitions to be used for processing the document. The definitions identify pieces of information on the document by, for example, absolute location, variable location, or relative location, or by proximity to key words and/or symbols. These definitions are described in further detail below by way of an illustrative example.

FIG. 9 illustrates a more detailed information processing data flow diagram for the process document module 2.3. The process document module 2.3 processes the document after the document has been stored in the system. The process document module 2.3 gathers the appropriate information which has been stored, and creates input file(s) 2.3.3 for the selected application unit. The process document module 2.3 then transmits the input file(s) via bus 240 and/or communication interface 4.0 to an application unit 270, an output device such as printer 260, or to main memory 250.

The process document module 2.3 includes an extract data module 2.3.1. This module extracts data off of the document in accordance with the user's instructions, for example, the user-defined template, or through the interactive mode.

The process document module 2.3 also includes a preapplication process module 2.3.2 which gathers and associates information extracted from the document in accordance with content instructions. This module 10 prompts the user for any additional information required to satisfy the relationships defined by the content instructions. The preapplication process module 2.3.2 also places the selected information into the transmission format defined by the transmission format instructions.

The preapplication process module 2.3.2 also generates the input file 2.3.3 for the selected application in accordance with the appropriate instructions. The input file 2.3.3 is then transmitted to bus 240 and/or communication interface 4.0 for transmission to a particular application unit 270.

FIG. 10 illustrates a detailed output data flow diagram for output module 3.0. Output module 3.0 outputs a textual and/or image copy of the document. In the 20 FIG. 1 preferred embodiment, output module 3.0 is implemented by printer 260, associated software, and associated interface circuitry.

Operation

Examples of operation of a preferred embodiment will now be described.

The user enters the system by providing instructions to the information processing module 2.0. The user then instructs the information processing module 2.0 to conduct maintain library processing, maintain definitions processing, or process document processing.

If the user selects maintain library processing, the user then provides instructions to maintain or modify the document library through the maintain library module 2.1. For example, the user can direct the inputting and storage of a hard copy document 100 or can retrieve and output a document. The user requests inputting of a document through the store document module 2.1.1. The system then prompts the user to specify a storage location for the inputted document. The document is then read in by the input process module 1.0. A textual copy and/or an image copy are stored into the document file 2.1.4. Errors which have occurred during inputting are identified and corrected by the correct errors module 2.1.2 and the user. The corrections are reflected in the document information stored in document file 2.1.4.

The retrieve document module 2.1.3 is used to retrieve and output a document. The system prompts the user to specify the storage location of a document and the type of document copy, for example, a textual or an image copy, to be outputted. The document is then outputted by the output process module 3.0.

If the user initially selected maintain definitions processing, the user would instruct the system to maintain and/or modify parameter definitions through the maintain definitions module 2.2. For example, the user can define and maintain a document template for extracting selected portions of information off of the hard copy document. The user can use the template to extract selected portions of information off of the hard copy document when the document is originally inputted, or alternatively, the user can use the template to identify

selected portions of information for extraction off of an image copy of the document. In creating the template, the user identifies pieces of information on the document to be extracted and assigns a variable name, or subfile, to each piece of data.

The location of data to be extracted can be defined in a number of ways other than by use of a template. For example, the user can designate the absolute location of information on the document with respect to a grid overlaid on the document, e.g., always on line 3, starting in column 1. The user can also identify information by specifying the relative location of information to be extracted, e.g., always two lines below the piece of data named "salutation" starting in column 3. The user can also specify the location of information to be extracted by variable location specification. For example, if the hard copy document is a letter, the module would conduct a key word search for the term "Dear Sir:". Whenever this term "Dear Sir:" is located, this piece of data would be associated with the variable specified by the user, for example, the variable "salutation." In addition, a defined set of conventional symbols can be used to signify certain recurring data items for the convenience of users of the instant invention. For example, a "@" symbol can be used to delineate the vendor name as follows: "@XYZ Corporation@". Other examples of the use of symbols to delineate information will be described with reference to FIG. 14.

The maintain definition module 2.2 is also used to maintain data relationships in accordance with content instructions and to maintain input file formats in accordance with transmission format instructions. Relationships are defined and maintained between pieces of data, specified by, for example, the names of variables, through the define relationships module 2.2.2. The names of pieces of data on the document are retrieved by, for example, the define template module 2.2.1, and are passed to the define relationships module 2.2.2. The user may then provide any additional pieces of data needed to generate an input file for a particular application program or unit, such as an input file line number. The user, the applications software, and/or instructions previously stored in memory then establishes the contents of the input file by defining relationships between pieces of data using content instructions. Specific examples of content instructions will be discussed below in conjunction with FIGS. 11, 12A, 12B, 12C, 13A, 13B, and 13C.

The user and/or the applications software defines and maintains the transmission format of the input file to be used by a particular application program or unit through the define format module 2.2.3 in accordance with transmission format instructions. This is accomplished by defining the parameters to be used by the preapplication process module 2.3.2 in generating an input file. Parameters which would typically be required to generate an input file would include the character type, e.g., text or pixel; delimiters used between pieces of data, e.g., a slash or a semicolon; end of line characters, e.g., a carriage return or a line feed; and end of file characters. Examples of transmission formats will be described in further detail below in conjunction with FIGS. 11, 12A, 12B, 12C, 13A, 13B, and 13C.

If the user initially selected process document processing, the interface will then proceed to process the document through use of the process document module 2.3. For example, the user can extract specific portions of data from an image copy of a document, can generate

an input file for transmission to an application program, or can directly process information interactively with an application program.

If the user desires to extract specific portions of data from an image copy of a hard copy document which has already been stored in memory, the user uses the extract data module 2.3.1 to identify a document to be processed. The document is then retrieved by the retrieve document module 2.1.3 and passed to the extract data module 2.3.1. The user can also select parameter definitions through the select definitions module 2.2.4.

The selected document template or parameter definition is passed to the extract data module 2.3.1. The extract data module 2.3.1 extracts pieces of data from the image copy of the document, as defined by the document template definition or the parameter definitions or both. This document data is then passed to preapplication process module 2.3.2.

The interface generates input file(s) 2.3.3 by use of the preapplication process module 2.3.2. The selected data relationship definition, as defined by the content instructions, and the selected record format definitions, as defined by the transmission format instructions, are passed to the preapplication process module 2.3.2. The preapplication process module 2.3.2 assembles the input file in accordance with the content instructions. The preapplication process module 2.3.2 also prompts the user for any additional pieces of data which need to be provided by the user. The input file is converted to the desired transmission format in accordance with the transmission format instructions. This physically formatted data is then stored in the input file 2.3.3.

The user can also use an application program to process information by loading the particular application program into the computer 230 rather than by sending the input file to a remote application unit 270.

An illustrative example of the processing described above will now be described. The user inputs instructions via keyboard 234 or another input device which indicate that the user desires to input and store a document. The computer 230 then prompts the user for the name of the document. In this example, the user desires to input the document of FIG. 2 and therefore names the document "XYZ Corp. Bill 12/01/86." The computer then prompts the user to feed the hard copy document 100 into the scanner 210. The image of the hard copy document is displayed on display 232. The computer then prompts the user to identify the account number on the document. By use of the mouse 236 or other input device to position a cursor on the display, the user indicates the location of the account number. The account number is then read-in to a subfile named "Account Number." This process proceeds until all of the desired information has been read-in and stored.

In this particular example, no errors were encountered while inputting the document. The user then directs that the document be stored for future reference in a document file.

Some time later, the user desires to retrieve and output the document and to generate input files based on information from the document. The computer 230 prompts the user for the name of the document and the type of output. The user responds with "XYZ Corp. Bill 12/01/86" for a printed textual copy. The document is then retrieved from the document file and passed to the printer 260 for printing.

In order to generate an input file for a specific application program, the user selects the option to define a

document template for use when each month's XYZ Corporation bill arrives. Accordingly, the user instructs the system to display a copy of an XYZ Corporation bill on the display 232. The user then identifies pieces of data by absolute locations. That is, the user assigns specific names to information located at specific portions of the document. In this example, the user would input the following information:

Vendor-text, line 1, one line, column 1, 80 characters;
 Account number-numeric, line 6, one line, column 25, 9 characters;
 Statement date-date, line 9, one line, column 25, 8 characters;
 Payment date-date, line 11, one line, column 25, 8 characters;
 Previous balance-currency, line 7, one line, column 75, 9 characters;
 New charges-currency, line 8, one line, column 75, 9 characters;
 Other debits-currency, line 10, one line, column 75, 9 characters;
 Finance charges-currency, line 12, one line, column 75, 9 characters;
 Payments-currency, line 13, one line, column 75, 9 characters;
 Other credits-currency, line 14, one line, column 75, 9 characters;
 New balance-currency, line 15, one line, column 75, 9 characters.

The user also identifies data with variable locations. In this particular example, a variable location is specified as follows:

Heading 2-line, value="Mail To:"

The identification of Heading 2 as line information means that the system will search for occurrences of the character string "Mail To:" and assign the line number which contains this character string to Heading 2.

The user also identifies data by relative locations. In this example, the user identifies the following relative location:

Mail To-text, Heading 2+1, 3 lines, column 60, 25 characters per line.

The instructions above instruct the system to assign the textual information beginning on one line after Heading 2 and continuing for 3 lines, in column 60, to the Mail To subfile.

As an alternative to inputting the actual line, column, and character numbers, the user can identify desired portions of the document by blocking, or highlighting, the desired portions using the mouse or other input device. In this case, the computer converts the highlighted portions into corresponding line, column, and character numbers.

FIG. 11 lists data corresponding to the hard copy document of FIG. 2 and the associated variable or subfile names.

Next, the user desires to define data relationships in accordance with content instructions. Examples of the type of contents which can be specified by a user are illustrated in FIGS. 12A, 12B, and 12C.

In this particular example, three separate departments of ABC Corporation require information from the XYZ Corporation bill. The first department requires vendor, account number, statement date, payment date, previous balance, new charges, debits, finance charges, payments, and new balance information. The second and third departments require mail to information and previous balance information. Each of these departments

have their own application program which utilizes this information.

The user employs content instructions to designate how pieces of information, which have been extracted off of hard copy document 100, are directed to particular departments, that is, particular application programs. FIG. 12A illustrates the contents of the information to be transmitted to the first department. FIG. 12B illustrates the information to be transmitted to the second department. FIG. 12C illustrates the information to be transmitted to the third department. The content instructions, therefore, parse the information shown in FIG. 11 to various application programs, as shown by FIGS. 12A, 12B, and 12C. Content instructions can also be used to identify additional pieces of data which are required for the input files of the particular application programs. In this particular example, the specific application programs from the three departments all require numeric record number information, numeric horizontal position information, numeric vertical position information, and date received information. The horizontal and vertical position information is used by the application program to specify the location of the received information on a spreadsheet application program, in this example. The user may know in advance the content format required by each application program, that is, in this example, the location and type of information specified on the spreadsheet. The user may also employ the split display mode described with reference to FIG. 3B to generate content format instructions.

Using the content instructions, the user establishes the following contents for the input file corresponding to FIG. 12A:

Record number, horizontal position, vertical position, vendor;
 Record number, horizontal position, vertical position, account number;
 Record number, horizontal position, vertical position, statement date;
 Record number, horizontal position, vertical position, date received;
 Record number, horizontal position, vertical position, payment date;
 Record number, horizontal position, vertical position, previous balance;
 Record number, horizontal position, vertical position, new charges;
 Record number, horizontal position, vertical position, finance charges;
 Record number, horizontal position, vertical position, payments;
 Record number, horizontal position, vertical position, new balance.

Next, transmission format instructions are employed to define the transmission format of the input file for a specific application program or unit. FIG. 13A illustrates the transmission input file corresponding to FIG. 12A. FIG. 13B illustrates the transmission input file corresponding to FIG. 12B. FIG. 13C illustrates the transmission input file corresponding to FIG. 12C. A comparison of FIGS. 12B and 12C reveals that FIGS. 12B and 12C have the same contents. However, the information illustrated in FIG. 12B is being sent to a different application program than the information in FIG. 12C. These application programs require different transmission input formats, as illustrated in FIGS. 13B and 13C. More specifically, the application program that receives the input file illustrated in FIG. 13B uses

the greater than sign as a delimiter whereas the application program which receives the transmission input file shown in FIG. 13C uses a back-slash as the delimiter.

After the contents and the transmission format for the input file have been defined, and any additional information has been inputted, the input file is assembled and transmitted to the particular application program.

FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention. The hard copy document illustrated in FIG. 14 is first scanned and information from the hard copy document is stored into a memory. The interface 200 then identifies portions of the hard copy document corresponding to various variables by recognizing a defined set of symbols. In the FIG. 14 example, triangles delineate the mailing address, circles delineate the statement date, and squares delineate the new charges. Information from these portions of the hard copy document is stored in the corresponding memory locations or subfiles for each variable. The same set of symbols can be used to identify the same information from one document to the next. Thus, even if the physical formats of documents are not fixed from one document to the next, a diversity of hard copy documents can be processed without manually inputting data by 25 recognition of the defined symbols.

Examples of readily available application programs are Quicken and Lotus 1,2,3 both of which are widely utilized in the business community. Quicken, for example is an easy-to-utilize program for writing checks and preparing business records. Payee, amount and address information may readily be transmitted from scanner memory 220 and/or main memory 250 to the Quicken application program for check writing functions and ledger keeping purposes. Lotus is a well known spreadsheet program which may process data input into specified cells once this data is placed in conventional Lotus format.

Thus, the instant invention provides an integrated and comprehensive system for handling information from a hard copy document, thus permitting a paperless office. In addition, the invention permits data, extracted off of a hard copy document, to be easily manipulated into various logical and transmission formats required by a particular application unit. The invention also 45 provides a low cost system for inputting information from a wide variety of hard copy documents into a memory.

The foregoing description has been set forth merely to illustrate preferred embodiments of the invention and 50 is not intended to be limiting. Modifications are possible without departing from the scope of the invention.

For example, letters, checks, forms, pictures, reports, music scores, film, and other types of hard copy documents can be processed by the invention for accounts payable/receivable accounting, inventory control, record keeping, budgeting, data base management, music transcription, forms processing, computerized art, survey and questionnaire processing, statistical data analysis, correspondence processing and other applications.

Other automated digitizing units can be used in addition to or as an alternative to use of the scanner 210 as an input unit. Any electrical, magnetic, or optical device which extracts information off of a hard copy document, thereby eliminating the need to manually input significant amounts of information from the hard copy document is suitable for use as an automated digitizing unit. In addition, information can be input by user re-

sponses and digital and analog signals generated from various devices, and from computer files from other computer systems. Suitable hardware for inputting data includes a keyboard, a light pen, a mouse, a touch screen, a laser scanner, a microphone, a tablet, a disk drive, a magnetic tape drive, and a modem.

The interface 200 can also output information in forms other than a hard copy of textual or image information. For example, the interface 200 can output system responses, computer files, and digital and analog signals for transmission to other computer systems or to control systems. Suitable hardware for outputting information includes a disk drive, a magnetic tape drive, a cathode ray tube, a plasma screen, a printer, a plotter, a film developer, an amplifier, and a modem.

Since modifications of the described embodiments incorporating the spirit and substance of the invention may occur to persons skilled in the art, the scope of the invention should be limited solely with respect to the appended claims and equivalents.

What is claimed is:

1. A method of processing information from a diversity of hard copy documents, said method comprising the steps of:

(a) receiving output representing a diversity of hard copy documents from an automated digitizing unit and storing information from said diversity of hard copy documents into a memory, said information not fixed from one document to the next;

(b) identifying portions of said hard copy documents corresponding to a first variable; and

(c) storing information from said portions of said hard copy documents corresponding to said first variable into memory locations for said first variable.

2. A method as set forth in claim 1, wherein step (b) includes displaying an image of a hard copy document on a display based on the contents of said memory.

3. A method as set forth in claim 1, further comprising the steps of:

identifying portions of said hard copy documents corresponding to a second variable, said portions of said hard copy documents corresponding to said second variable being different from said portions of said hard copy documents corresponding to said first variable; and

storing information from said portions of said hard copy documents corresponding to said second variable into memory locations for said second variable.

4. A method as set forth in claim 1, wherein step (b) includes prompting identification of said portions of said hard copy documents corresponding to said first variable.

5. A method as set forth in claim 1, wherein step (c) includes storing image information from said portions of said hard copy documents corresponding to said first variable into said memory locations for said first variable.

6. A method as set forth in claim 1, wherein step (c) includes storing textual information from said portions of said hard copy documents corresponding to said first variable into said memory locations for said first variable.

7. A method as set forth in claim 1, further comprising the steps of detecting and correcting errors resulting from said inputting.

8. A method as set forth in claim 1, further comprising the step of utilizing a template to associate portions of said hard copy documents with specific variables.

9. A method as set forth in claim 1, further comprising receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a location on said hard copy documents containing a value of a specific variable.

10. A method as set forth in claim 1, further comprising receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a relative location on said hard copy documents containing a value of a specific variable.

11. A method as set forth in claim 2, further comprising the steps of:

identifying portions of said hard copy documents corresponding to a second variable, said portions of said hard copy documents corresponding to said second variable being different from said portions of said hard copy documents corresponding to said first variable; and

storing information from said portions of said hard copy documents corresponding to said second variable into memory locations for said second variable.

12. A method as set forth in claim 3, further comprising the step of storing image information from said portions of said hard copy documents corresponding to said second variable into said memory locations for said second variable.

13. A method as set forth in claim 3, further comprising the step of storing textual information from said portions of said hard copy documents corresponding to said second variable into said memory locations for said second variable.

14. A method as set forth in claim 3, further comprising the step of prompting identification of said portions of said hard copy documents corresponding to said second variable.

15. A method of processing information from a diversity of hard copy documents, said method comprising the steps of:

(a) scanning a diversity of hard copy documents and storing information from said diversity of hard copy documents into a memory, said information not fixed from one document to the next;

(b) identifying portions of said hard copy documents corresponding to a first variable; and

(c) storing information from said portions of said hard copy documents corresponding to said first variable into memory locations for said first variable.

16. A method as set forth in claim 15, wherein step (b) includes displaying an image of a hard copy document on a display based on the contents of said memory.

17. A method as set forth in claim 15, further comprising the steps of:

identifying portions of said hard copy documents corresponding to a second variable, said portions of said hard copy documents corresponding to said second variable being different from said portions of said hard copy documents corresponding to said first variable; and

storing information from said portions of said hard copy documents corresponding to said second variable into memory locations for said second variable.

18. A method as set forth in claim 15, wherein step (b) includes prompting identification of said portions of

said hard copy documents corresponding to said first variable.

19. A method as set forth in claim 15, wherein step (c) includes storing image information from said portions of said hard copy documents corresponding to said first variable into said memory locations for said first variable.

20. A method as set forth in claim 15, wherein step (c) includes storing textual information from said portions of said hard copy documents corresponding to said first variable into said memory locations for said first variable.

21. A method as set forth in claim 15, further comprising the steps of detecting and correcting errors resulting from said scanning.

22. A method as set forth in claim 15, further comprising the step of utilizing a template to associate portions of said hard copy documents with specific variables.

23. A method as set forth in claim 15, further comprising receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a location on said hard copy documents containing a value of a specific variable.

24. A method as set forth in claim 15, further comprising receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a relative location on said hard copy documents containing a value of a specific variable.

25. A method as set forth in claim 16, further comprising the steps of:

identifying portions of said hard copy documents corresponding to a second variable, said portions of said hard copy documents corresponding to said second variable being different from said portions of said hard copy documents corresponding to said first variable; and

storing information from said portions of said hard copy documents corresponding to said second variable into memory locations for said second variable.

26. A method as set forth in claim 17, further comprising the step of storing image information from said portions of said hard copy documents corresponding to said second variable into said memory locations for said second variable.

27. A method as set forth in claim 17, further comprising the step of storing textual information from said portions of said hard copy documents corresponding to said second variable into said memory locations for said second variable.

28. A method as set forth in claim 17, further comprising the step of prompting identification of said portions of said hard copy documents corresponding to said second variable.

29. A method of processing data extracted from a diversity of hard copy documents, said method comprising the steps of:

(a) receiving output representing a diversity of hard copy documents from an automated digitizing unit and storing information from said diversity of hard copy documents into a memory as stored document information;

(b) selecting portions of said stored document information in accordance with content instructions defining portions of said stored document information required by an application unit;

(c) formatting selected stored document information into a transmission format used by said application unit based on transmission format instructions; and
 (d) transmitting formatted selected stored document information to said application unit.

30. A method as set forth in claim 29, wherein step (a) includes storing textual information representing characters on said hard copy documents.

31. A method as set forth in claim 29, wherein step (a) includes storing digitized image information representing the actual appearance of said hard copy documents. 10

32. A method as set forth in claim 29, further comprising detecting and correcting errors in said stored document information resulting from said inputting.

33. A method as set forth in claim 29, wherein step (a) includes the step of utilizing a template to associate portions of said hard copy documents with specific variables. 15

34. A method as set forth in claim 29, wherein step (a) includes receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a location on said hard copy documents containing a value of a specific variable. 20

35. A method as set forth in claim 29, wherein step (a) includes receiving instructions identifying at least one character or symbol, located on said hard copy documents, which identifies a relative location on said hard copy documents containing a value of a specific variable. 25

36. A method as set forth in claim 29, further comprising the step of printing textual copies of said hard copy documents based on said stored document information. 30

37. A method as set forth in claim 29, wherein step (a) includes receiving output representing a diversity of 35 hard copy documents from a scanner.

38. A method as set forth in claim 31, further comprising the step of printing copies of said hard copy documents based on said digitized image information.

39. An application program interface, comprising:
 an automated digitizing unit which extracts information from a diversity of hard copy documents and stores said information from said diversity of hard

copy documents in a memory as stored document information;

a processor selecting portions of said stored document information in accordance with content instructions, said content instructions designating portions of said stored document information required by an application unit;

a formatter formatting selected stored document information into a transmission format used by said application unit based on transmission format instructions; and

an output unit transmitting formatted selected stored document information to said application unit.

40. An interface as set forth in claim 39, wherein said stored document information includes textual information representing characters on said hard copy documents.

41. An interface as set forth in claim 39, wherein said stored document information includes digitized image information representing the actual appearance of said hard copy documents.

42. An interface as set forth in claim 39, further comprising an error correcting unit detecting and correcting errors resulting from extracting by said automated digitizing unit.

43. An interface as set forth in claim 39, further comprising:

a template definition unit for defining a template which associates locations on said hard copy documents with specific variables.

44. An interface as set forth in claim 39, further comprising a search unit for searching for at least one character or symbol, located on said hard copy documents, which identifies a location on said hard copy documents containing a value for a specific variable.

45. An interface as set forth in claim 39, wherein said automated digitizing unit includes a scanner.

46. An interface as set forth in claim 41, further comprising a printer which prints out copies of the actual appearance of said hard copy documents based on said digitized image information.

* * * * *

United States Patent

[19]

[11]

Patent Number:

5,625,465

Lech et al.

[45] Date of Patent: Apr. 29, 1997

[54] INFORMATION PROCESSING
METHODOLOGY

[75] Inventors: Robert Lech, Jackson; Mitchell A. Medina, Essex Fells; Catherine B. Elias, Plainsboro, all of N.J.

[73] Assignee: International Patent Holdings Ltd., Hamilton, Bermuda

[21] Appl. No.: 348,224

[22] Filed: Nov. 28, 1994

Related U.S. Application Data

[63] Continuation of Ser. No. 143,135, Oct. 29, 1993, Pat. No. 5,369,508, which is a continuation of Ser. No. 672,865, Mar. 20, 1991, Pat. No. 5,258,855.

[51] Int. Cl.⁶ H04N 1/40

[52] U.S. Cl. 358/448; 358/403; 358/406; 382/209; 382/175; 382/180; 395/768

[58] Field of Search 382/209, 175, 382/180, 317; 358/462, 448, 400, 401, 403, 447, 449, 451, 452, 453, 460, 463, 467, 470, 471, 474

[56] References Cited

U.S. PATENT DOCUMENTS

4,034,343	7/1977	Wilmer	340/146.3 MA
4,667,248	5/1987	Kanno	358/280
4,760,606	7/1988	Lesnick et al.	382/48
5,034,990	7/1991	Klees	382/22
5,095,445	3/1992	Sekiguchi	364/514
5,140,650	8/1992	Casey et al.	382/61
5,148,520	9/1992	Morgan	395/148

5,153,927	10/1992	Yamanari	382/61
5,228,100	7/1993	Takeda et al.	382/175
5,258,855	11/1993	Lech et al.	358/462
5,367,619	11/1994	Dipaolo et al.	395/149

OTHER PUBLICATIONS

"Kurzweil 5200 Intelligent Scanning System", Xerox Imaging Systems, Inc., 1990.
Que's Computer User's Dictionary, 2nd Ed., Bryan Pfaffenberger (author); 1991; p. 144.

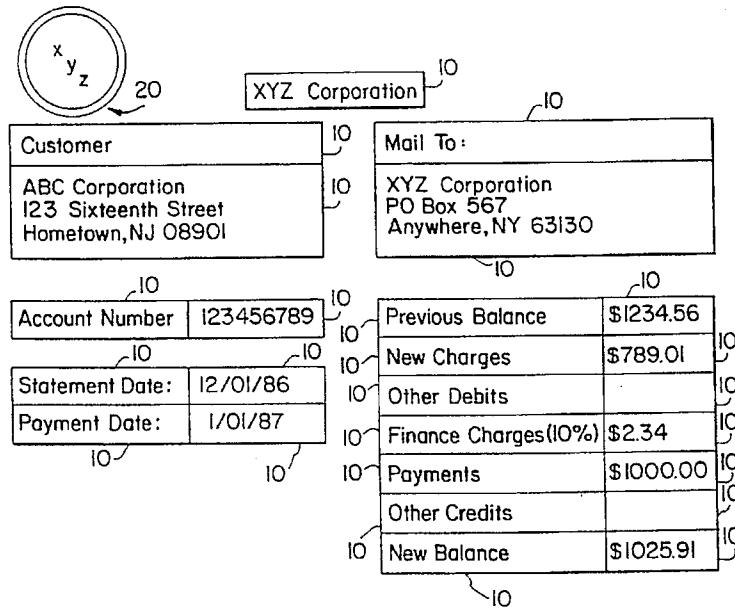
Primary Examiner—Joseph Mancuso
Assistant Examiner—Jerome Grant, II
Attorney, Agent, or Firm—Foley & Lardner

[57]

ABSTRACT

An information processing methodology gives rise to an application program interface which includes an automated digitizing unit, such as a scanner, which inputs information from a diversity of hard copy documents and stores information from the hard copy documents into a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which designate portions of the stored document information required by a particular application program. The selected stored document information is then placed into the transmission format required by a particular application program in accordance with transmission format instructions. After the information has been transmission formatted, the information is transmitted to the application program. In one operational mode, the interface interactively prompts the user to identify, on a display, portions of the hard copy documents containing information used in application programs or for storage.

53 Claims, 11 Drawing Sheets



Winter Sale In Effect Throughout January!

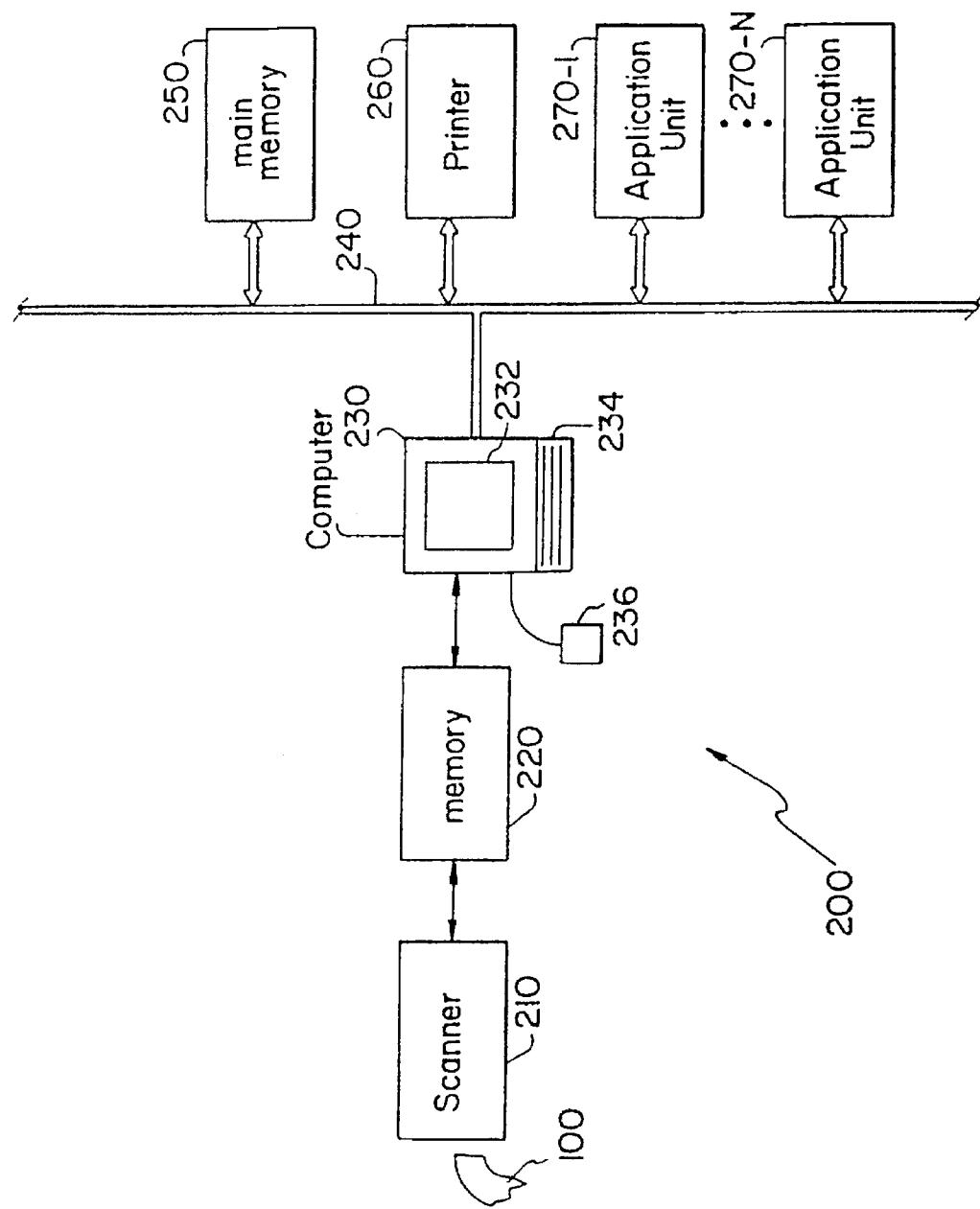
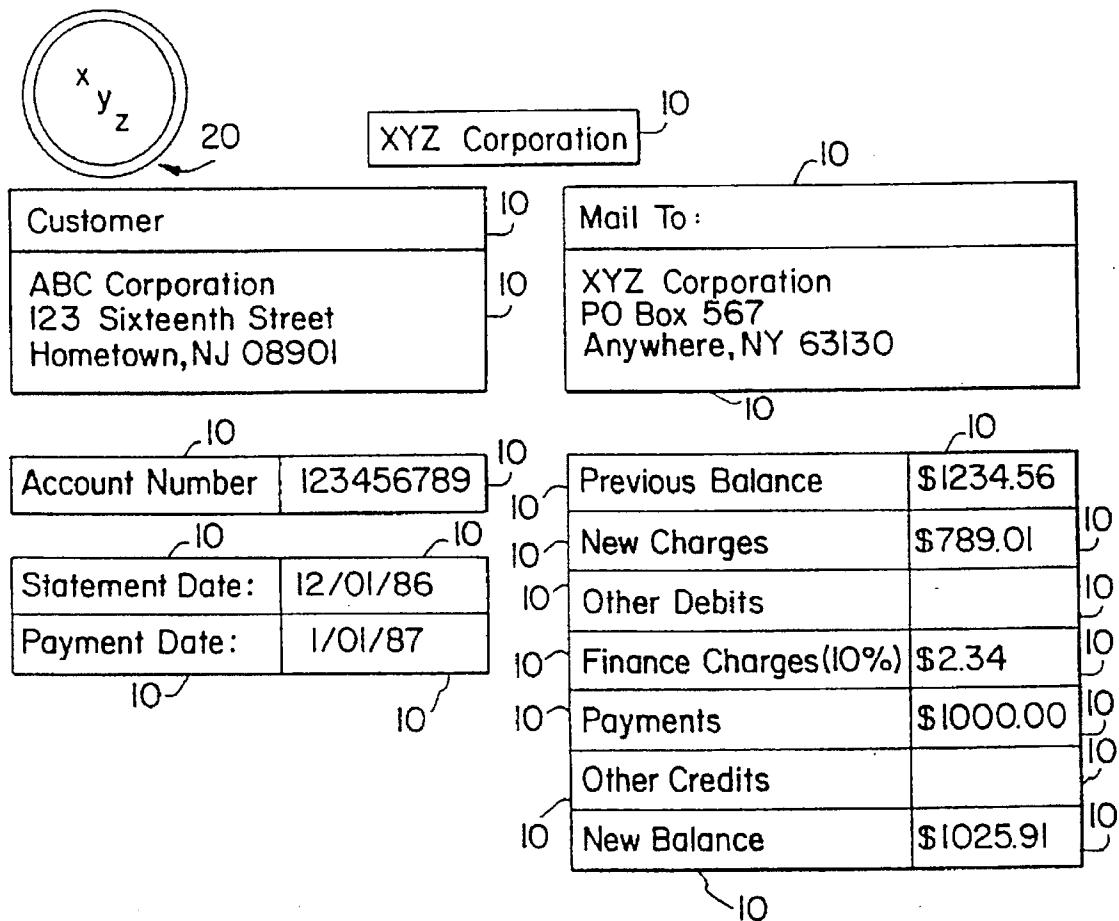


FIG. 1



Winter Sale In Effect Throughout January!

30

FIG. 2

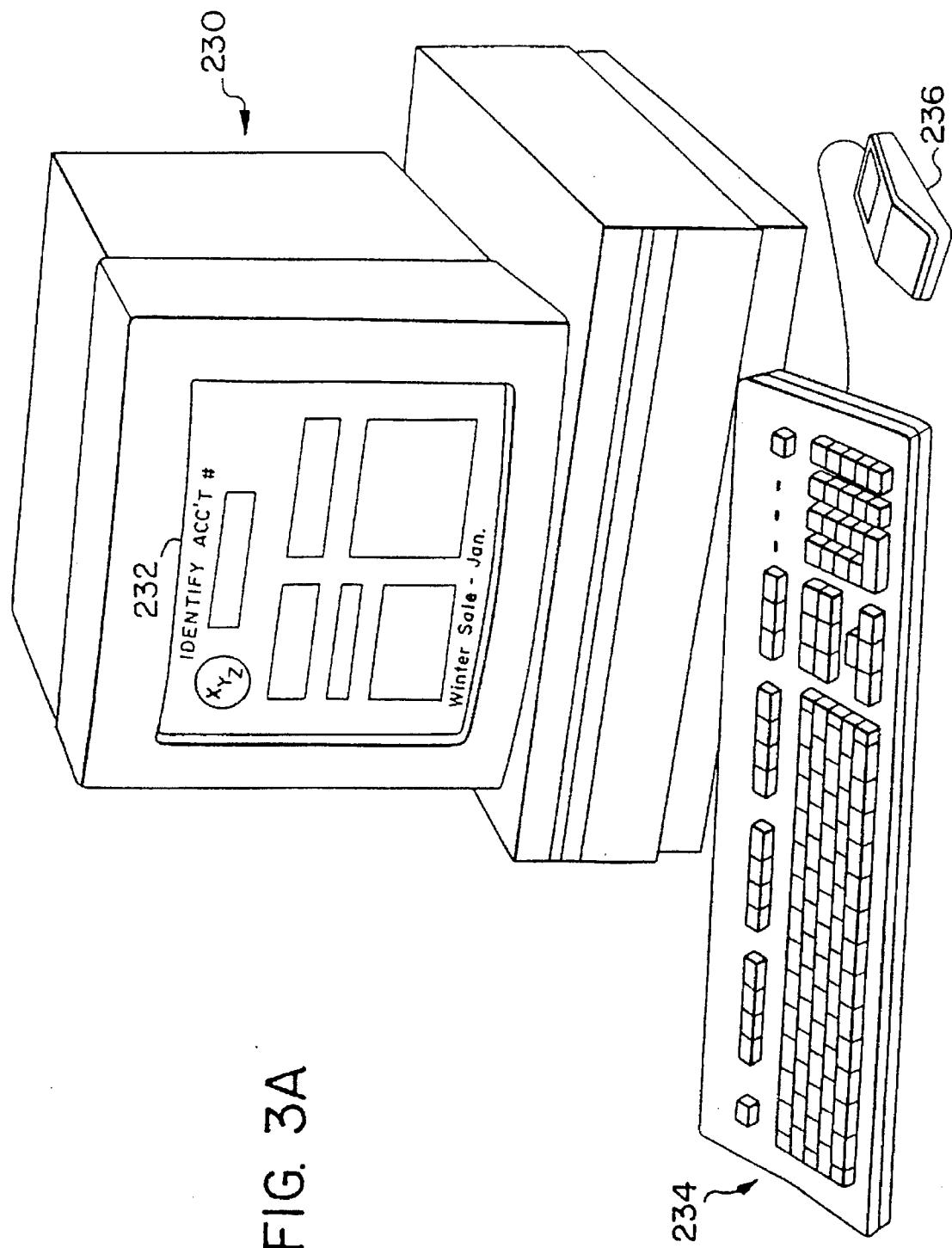


FIG. 3A

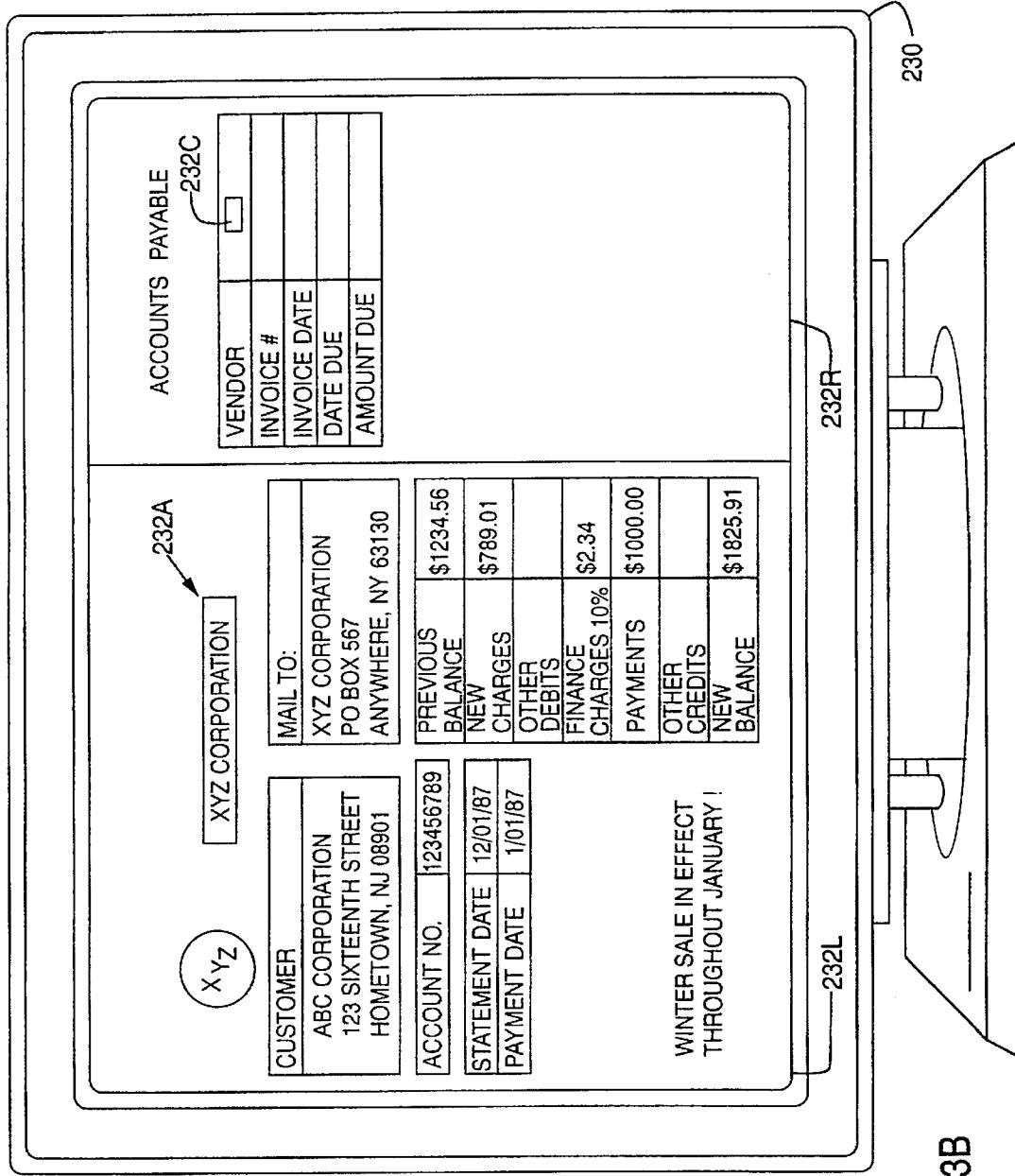


FIG. 3B

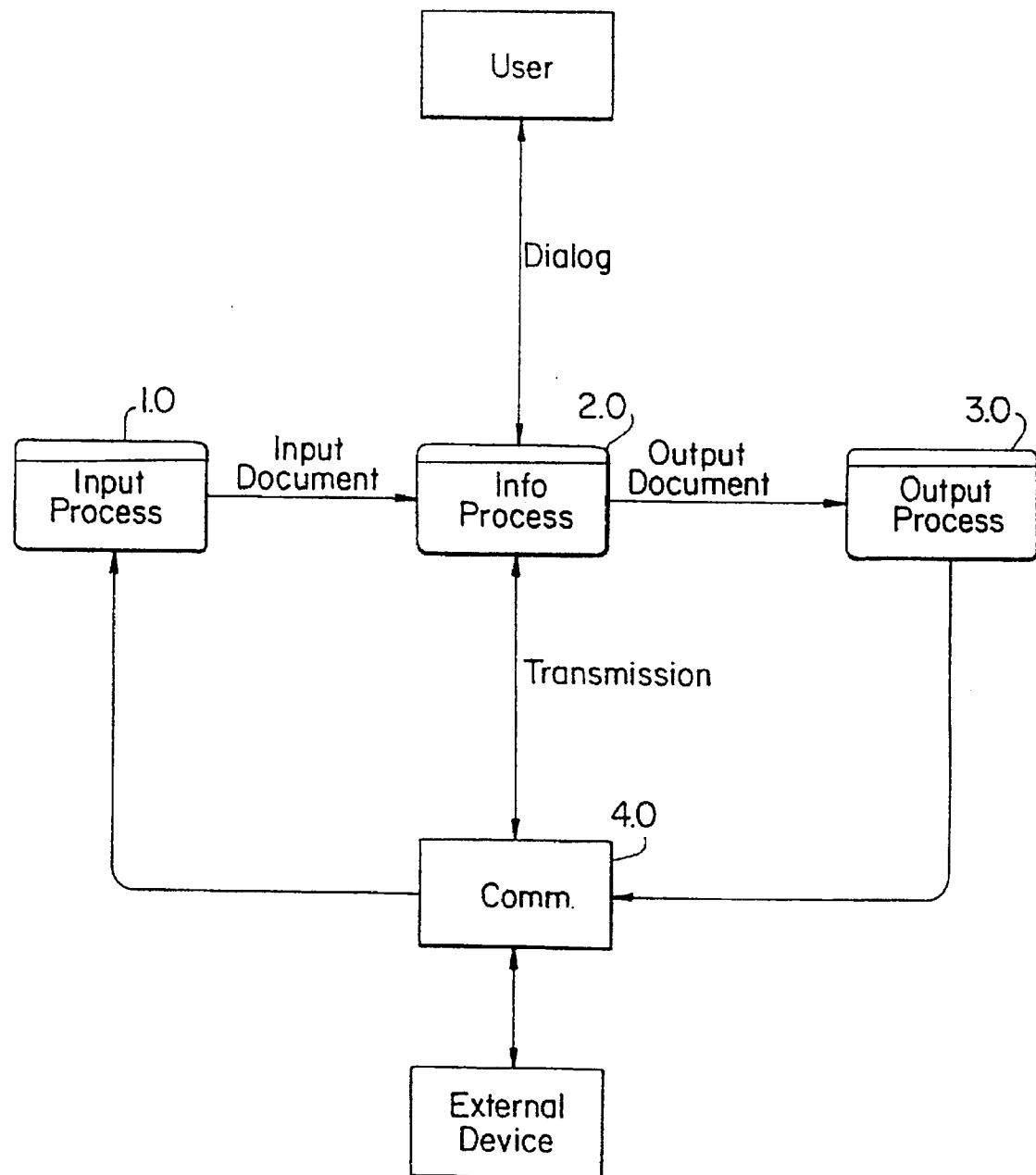


FIG. 4

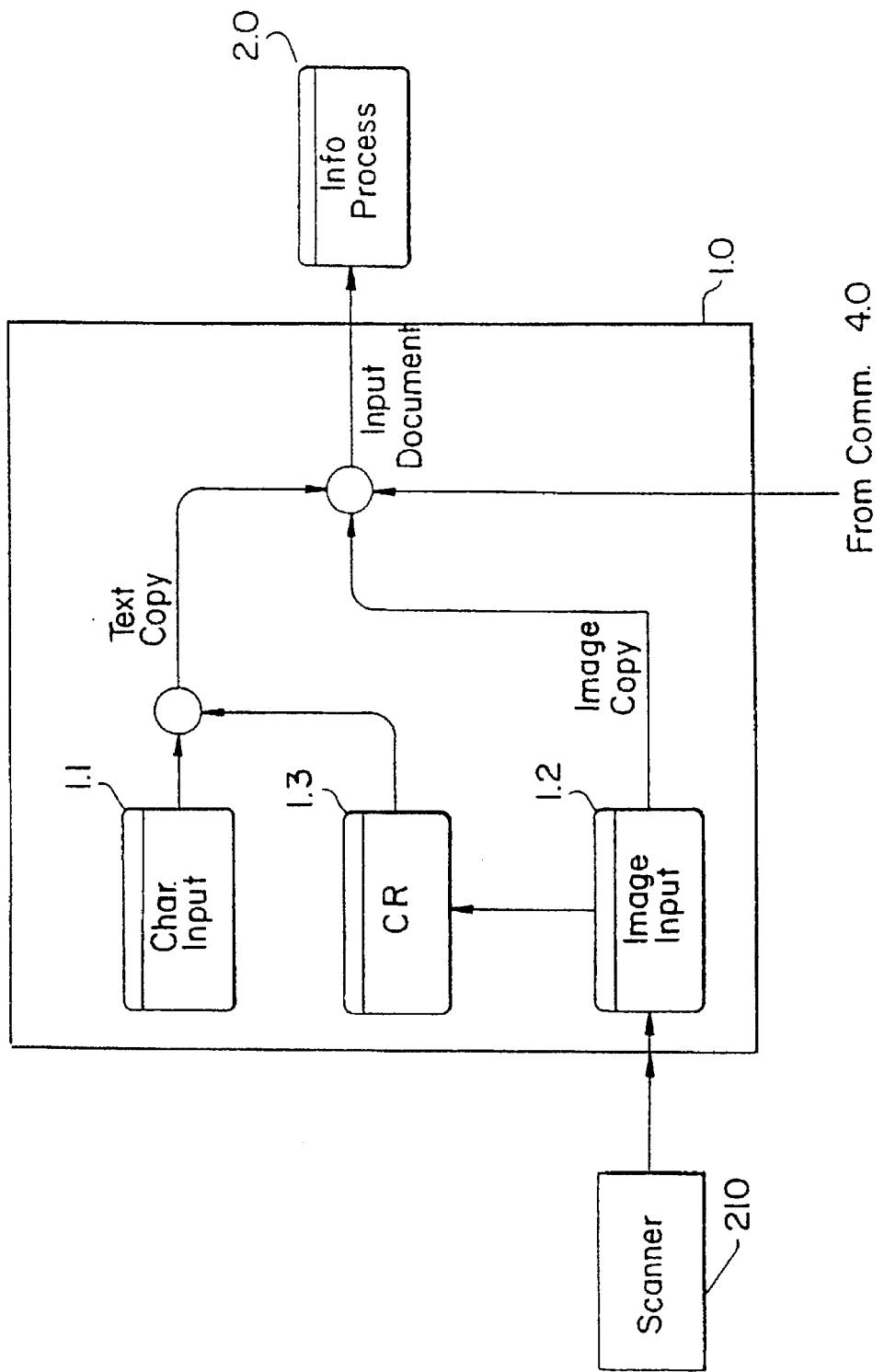


FIG. 5

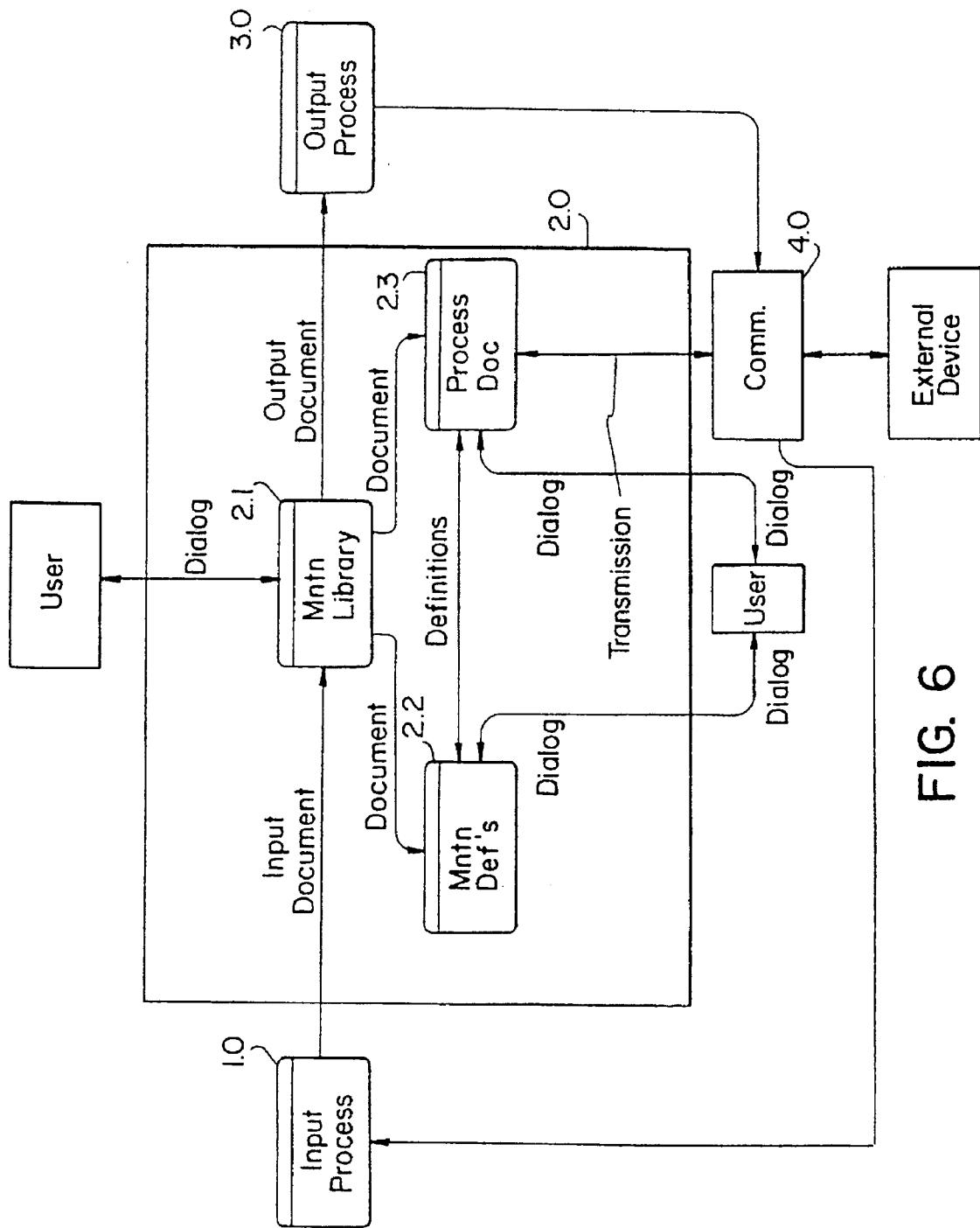


FIG. 6

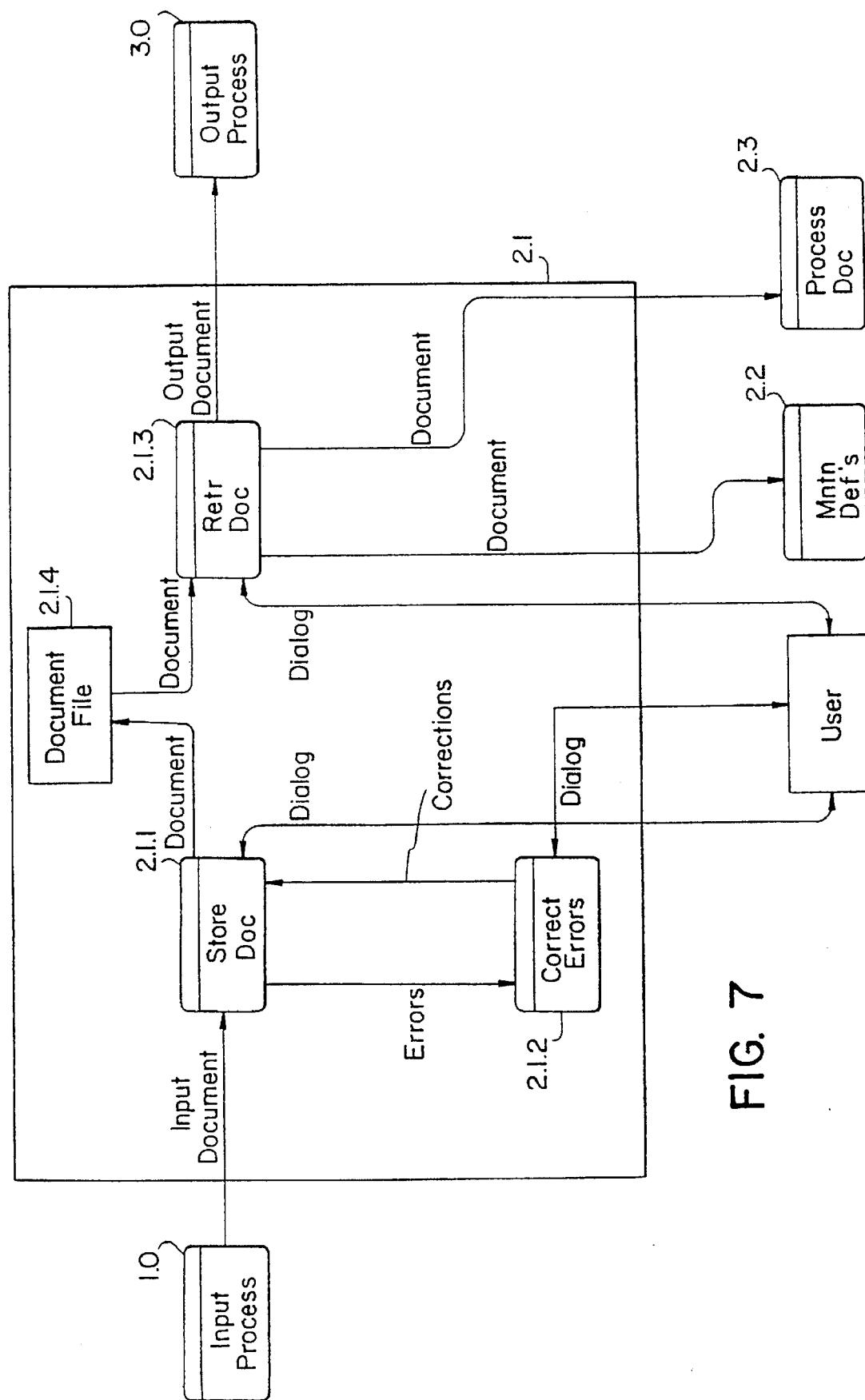


FIG. 7

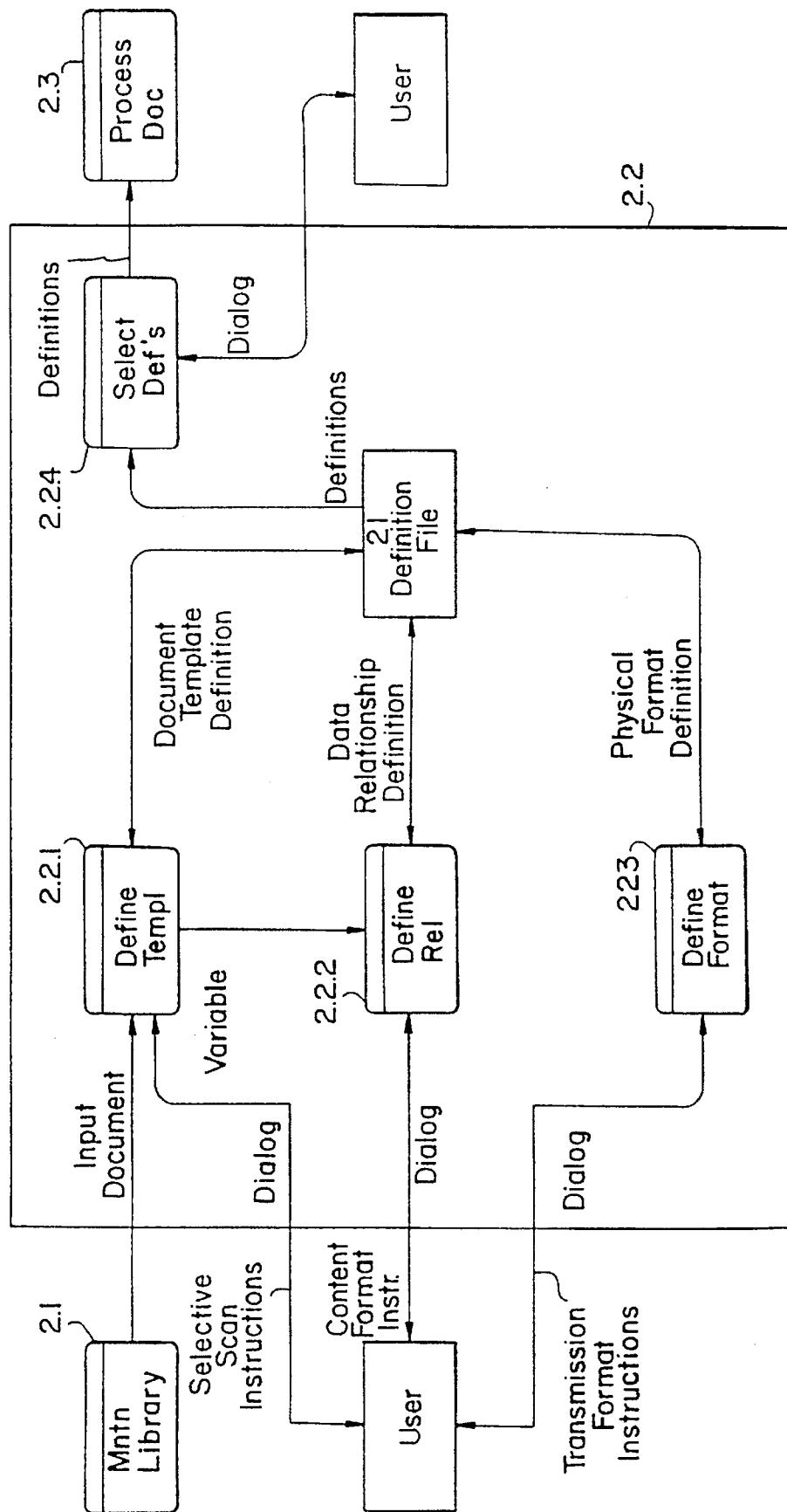


FIG. 8

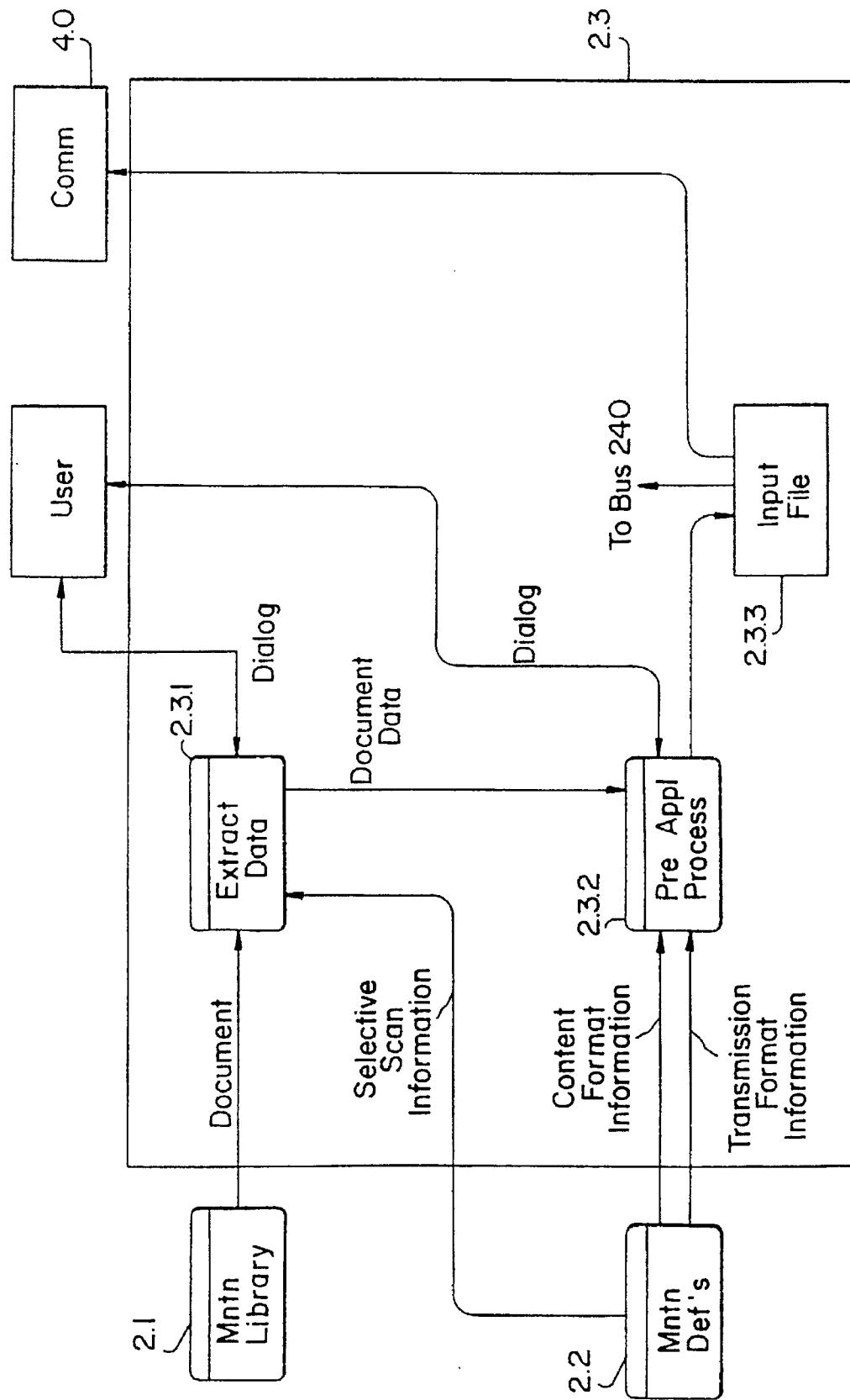


FIG. 9

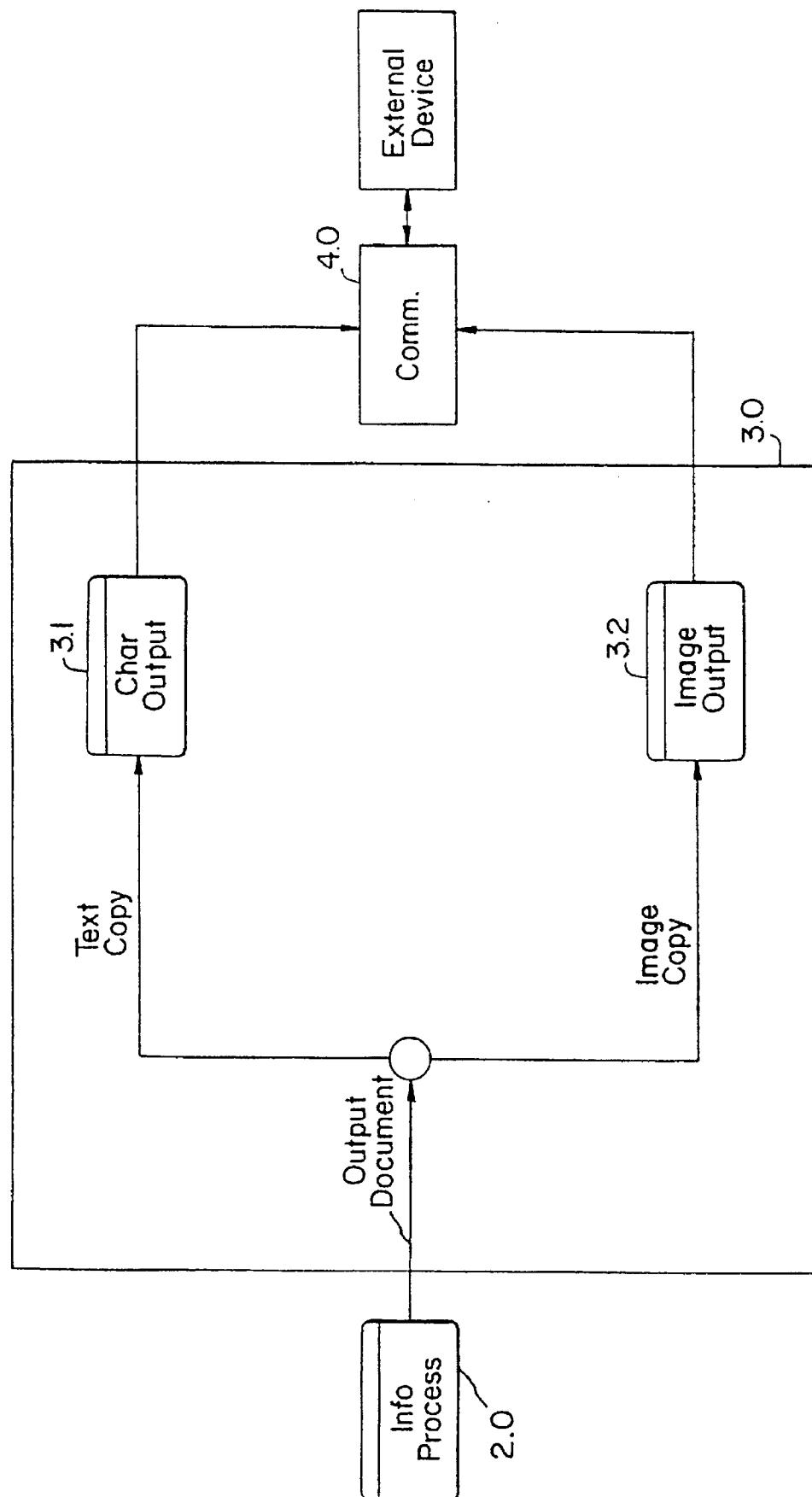


FIG. 10

INFORMATION PROCESSING METHODOLOGY

This application is a continuation of application Ser. No. 08/143,135, filed Oct. 29, 1993, now U.S. Pat. No. 5,369,508 which is a continuation of application Ser. No. 07/672,865, filed Mar. 20, 1991, now U.S. Pat. No. 5,258,855, issued Nov. 2, 1993.

BACKGROUND OF THE INVENTION

The invention is directed to a system for efficiently processing information originating from hard copy documents. More specifically, the invention is directed to a hard copy document to application program interface which minimizes the need to manually process hard copy documents.

In the past, information contained on hard copy documents was manually entered into a computer via the input controller of a particular computer. The original document was then filed away for future reference. Automatic input of data was limited to the input of Magnetic Ink Character Recognition (MICR) data and to Optical Character Recognition (OCR) data. This fixed-position data was forwarded directly to a dedicated computer application specifically designed to accommodate the input format. In more recent years, typewritten text has been mechanically inputted into a computer via a text file. Examples of this latter type of system are word processors and photo-typesetters.

These conventional systems have limitations which decrease the efficiency of processing information from a hard copy document. For example, the systems discussed above are limited in their application to MICR, OCR, or typewritten data. Parsing and processing data is limited to the particular requirements of the particular computer application which requires the input data. In addition, in these conventional systems, the actual hard copy document must be retained for future reference at great expense.

In a sophisticated computer network, different users may require different portions of the information contained on a hard copy document. For example, if the hard copy document is an invoice returned with payment of a bill, the accounting department may need all of the monetary information contained on the bill while the mailroom may need only customer address information, to update a customer's address. Therefore, there is a need for a system in which specific information from a hard copy document can be selectively distributed to various users.

Another problem with conventional systems is that users, even within the same company, may require that the information extracted from a hard copy document be transmitted to a particular application program in a specific transmission format. For example, one department in a company may use a particular application program which must receive information using a particular character as a delimiter and other departments may require the information in a different format using different delimiters.

Another problem, particularly for small businesses, is that current systems can not efficiently accommodate the inputting of information from a diversity of hard copy documents. A large business which receives many forms in the same format can afford a system which inputs a high volume of information in that format into memory. For example, it is cost-effective for a bank which processes hundreds of thousands of checks a month to buy a dedicated machine which can read information off of checks having a rigidly defined, or fixed, format. However, as the diversity of forms received

by a business increases relative to the number of forms that must be processed, it becomes less cost-effective to design a dedicated machine for processing each type of form format. This problem is particularly significant in small businesses which may, for example, receive fifty invoices a month, all in different, non-fixed, formats. It is frequently not cost-effective for a small business to design dedicated systems for inputting information in each of these various formats. This leaves a small business with no other practical alternative than to manually input the information off of each invoice each month.

SUMMARY OF THE INVENTION

It is an object of the invention, therefore, to provide an application program interface which allows a user to select specific portions of information extracted from a diversity of hard copy documents and allows the user to direct portions of this information to several different users in accordance with the needs of the particular user.

It is also an object of the invention to provide a cost-effective system for inputting hard copy documents which can accommodate hard copy documents in a diversity of formats.

It is another object of the invention to provide an application program interface which allows a user to put information, which is to be transmitted, into a particular transmission format, based upon the needs of the receiver of the information.

It is a further object of the invention to provide an application program interface which will allow the extraction, selection, formatting, routing, and storage of information from a hard copy document in a comprehensive manner such that the hard copy document itself need not be retained.

It is another object of the invention to provide a system which reduces the amount of manual labor required to process information originating from a hard copy document.

A further object of the invention is to reduce the time required to process information originating from a hard copy document so that a higher volume of transactions involving hard copy documents can be processed.

The invention provides an application program interface which inputs a diversity of hard copy documents using an automated digitizing unit and which stores information from the hard copy documents in a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which define portions of the stored document information required by a particular application unit. Selected stored document information is then formatted into the transmission format used by the particular application program based on transmission format instructions. The transmission formatted selected stored document information is then transmitted to the particular application program. The hard copy documents may contain textual information or image information or both.

The interface operates in three different modes.

In a first mode, the interface extracts all of the information from hard copy documents and stores this information in memory. Parsing of various portions of the extracted information is performed in accordance with content instructions.

In a second mode, the user operates interactively with the interface by use of a display and an input device, such as a mouse. In this second mode, a hard copy document is inputted and displayed on the display. The interface then

prompts the user to identify the location of various information. For example, the interface can ask the user to identify the location of address information on the hard copy document. In response, the user positions the mouse to identify address information using a cursor. The identified information is then stored as address information in memory. Subsequently, the interface again prompts the user to identify other pieces of information, which are then stored in the appropriate locations in memory. This process proceeds until all of the information which is desired to be extracted off of the hard copy document is stored in memory.

In a third mode of operation, selected portions of information are extracted off of hard copy documents in accordance with predetermined location information which has been specified by the user. For example, the user can define a template which specifies the location of information on hard copy documents. Templates can be formed in conjunction with second mode operation. Alternatively, the user can instruct the interface to search hard copy documents for a particular character or symbol, located on the hard copy documents. The information desired to be extracted off of the hard copy documents is specified relative to the location of this character or symbol.

The interface can also prompt or receive from an applications program or another information processing system, required information, content instructions, and format instructions.

Other objects, features, and advantages of the invention will be apparent from the following detailed description of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be described in further detail below with reference to the accompanying drawings, in which:

FIG. 1 illustrates hardware for implementing a preferred embodiment of the instant invention;

FIG. 2 illustrates an example of a hard copy document containing information to be processed by the instant invention;

FIGS. 3A and 3B are enlarged views of the computer of FIG. 1 used to explain how the invention interactively prompts a user to identify information;

FIG. 4 is an overall data flow diagram for the FIG. 1 preferred embodiment;

FIG. 5 is a detailed input data flow diagram for the FIG. 1 preferred embodiment;

FIG. 6 is a detailed information processing data flow diagram for the FIG. 1 preferred embodiment;

FIG. 7 is a more detailed information processing data flow diagram for the maintain library module of FIG. 6;

FIG. 8 is a more detailed information processing data flow diagram for the maintain definitions module of FIG. 6;

FIG. 9 is a more detailed information processing data flow diagram for the process document module of FIG. 6;

FIG. 10 is a detailed output data flow diagram for the FIG. 1 preferred embodiment;

Appendix 11 lists data corresponding to the hard copy document of FIG. 2;

Appendices 12A, 12B, and 12C illustrate examples of data which can be selected from the extracted data of FIG. 11 in accordance with content instructions.

Appendices 13A, 13B, and 13C illustrate examples of the data of FIGS. 12A, 12B, and 12C formatted in accordance with various transmission format instructions to form input files; and

Appendix 14 illustrates another example of a hard copy document containing information to be processed by the instant invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Hardware

The invention provides an interface between information originating from a hard copy document and a computer application unit which uses the information. The computer application unit can be a particular computer application program or a device which is controlled in accordance with instructions or information from the hard copy document.

15 The invention also allows storing a copy of the hard copy document in a memory and retrieving the copy of the hard copy document. By providing a comprehensive and integrated system which can accommodate almost all of the possible uses of information contained on a hard copy document, the instant invention allows for a paperless office.

The invention includes hardware and software necessary to extract, retrieve, and process information from the hard copy document. A copy of the actual image of the hard copy document is stored in memory. Textual information

25 extracted from the hard copy document is also stored in memory. Textual information is information, such as alphanumeric characters, which is recognized on the hard copy document and which is stored in a form which corresponds to the particular recognized character. For example, the extracted characters can be stored in the ASCII format in an electronic memory.

The user can have all of the information extracted from the hard copy document and stored in memory. Alternatively, the interface can interactively prompt the user 35 to identify specific pieces of information for storage. The interface can also extract specific pieces of information using a predefined template. The interface can also prompt or receive from another information processing system or an applications program desired information, content instructions, and format instructions.

The instant invention also provides for parsing information extracted from the hard copy document and for directing this parsed information to specific users or application programs as an input file.

45 The invention also permits the user to define the transmission format of the input file for a particular computer application unit.

FIG. 1 illustrates hardware for implementing a preferred embodiment of a hard copy document to application program interface according to the instant invention. The interface 200 processes information extracted off of hard copy document 100 and provides information to application units 270 in a form required by each particular application unit.

50 The interface extracts information off of a hard copy document 100 utilizing a scanner 210. The scanner 210 can be any type of scanner which extracts information off of hard copy documents, for example, an Optical Reader.

The scanned information is stored in a scanner memory 220 or in main memory 250, as will be described in greater detail below. If main memory 250 or another memory is available to store the scanned information, then scanner memory 220 can be omitted.

60 The information from scanner memory 220 or main memory 250 is transmitted to computer 230. In the preferred embodiment, computer 230 includes a display 232, a keyboard 234, and a mouse 236. The display 232 displays an

image of the hard copy document itself and/or information necessary to process the information extracted off of the hard copy document.

The computer 230 is used to select portions of the stored document information contained in memory in accordance with content instructions which define portions of the stored document information required by an application unit. These content instructions may be provided by the application program. Alternatively, the content instructions can be inputted via an input device such as a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

The computer 230 is also used to format selected stored document information into the transmission format used by an application unit based on transmission format instructions. The transmission format instructions may be provided by the application program. Alternatively, the transmission format instructions can be inputted via a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

Thus, the computer 230 is used to generate an input file for a particular application unit. The computer 230 is connected to scanner memory 220, main, or permanent, memory 250, a printer 260, and application units 270, via a bus 240. Although FIG. 1 illustrates use of a bus to connect components together, it is understood that any routing or connecting link, implemented in hardware or software or both, can be employed instead of, or in addition to, a bus. Instructions to or in the computer 230 control the main memory 250, the printer 260, the application units 270, and the bus 240. Instructions to or in computer 230 can also control exchanges of information with scanner memory 220.

When the computer 230 generates an input file for a particular document, the computer 230 can send this input file directly to an application unit or can store this input file in the main memory 250 until required by an application unit. The main memory 250 may also optionally store a copy of the image information for the hard copy document and the textual information for the hard copy document. Thus, the image information and textual information from the hard copy document can be retrieved and printed out on printer 260. In addition, image and textual information stored in scanner memory 220 or in main memory 250 can be used to form additional input files at the time of input or at a later time, based on content instructions and transmission format instructions. Thus, the invention can, at the discretion of the user, eliminate the need to retain copies of hard copy documents, permitting a paperless office.

The application units 270 include particular application programs and devices which are controlled in accordance with information contained on hard copy document 100.

FIG. 2 illustrates an example of a hard copy document 100 which contains information to be processed by the instant invention. The document illustrated in FIG. 2 is a bill from XYZ Corporation to customer ABC Corporation. FIG. 2 is only an example of a type of document that can be processed by the instant invention.

In a first operational mode, the scanner 210 stores all of the information extracted off of hard copy document 100 in the scanner memory 220 or, alternatively, in main memory 250. The extracted information is stored in two forms. The actual image of the hard copy document 100 is stored as image information in the scanner memory 220. In addition, the scanner memory 220 stores textual information recognized on the hard copy document 100 by, for example, employing standard character recognition software. In the preferred embodiment, the textual information is stored in

ASCII format. The scanner memory 220 can be, for example, an electronic, magnetic, or optical memory.

FIG. 3A illustrates an enlarged view of the computer 230 of FIG. 1. This view will be used to describe a second mode

5 of operation. In this second mode of operation, the hard copy document 100 is scanned and a copy of the document 100 is displayed on display 232 of computer 230, based on the contents of information temporarily stored in scanner memory 220. After the document is displayed on display 10 232, the computer 230 interactively prompts the user to identify the location of specific pieces of information on the hard copy document. In the FIG. 3A illustration, this prompt message is indicated as the message beginning with the arrow.

15 For example, the prompt message can ask the user to identify the location of account number information on the hard copy document. The user then uses an input device, such as keyboard 234 or mouse 236 or a touch screen, notepad, voice recognition device, or other input device to

20 position a cursor on the display to identify the location of the information requested by the prompt message. For example, the cursor could be used to define a block (which could be highlighted) containing the requested information, followed by a mouse "enter" click. In this example, the user would move the mouse to identify the location of the account number information contained on the hard copy document 100. The computer 230 then stores the information which has been identified by the user as account number information in the appropriate address or subfile or as the appropriate variable or parameter, or data field in memory. The computer then prompts the user to identify the location of other information on the hard copy document, such as, statement date information. The process proceeds until all of the desired information has been stored into the appropriate

25 30 35 locations in memory.

FIG. 3B illustrates a variation of the second mode for interactively prompting the user for information. In FIG. 3B, the display is split into two portions. A left-hand portion

40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995 1000 1005 1010 1015 1020 1025 1030 1035 1040 1045 1050 1055 1060 1065 1070 1075 1080 1085 1090 1095 1100 1105 1110 1115 1120 1125 1130 1135 1140 1145 1150 1155 1160 1165 1170 1175 1180 1185 1190 1195 1200 1205 1210 1215 1220 1225 1230 1235 1240 1245 1250 1255 1260 1265 1270 1275 1280 1285 1290 1295 1300 1305 1310 1315 1320 1325 1330 1335 1340 1345 1350 1355 1360 1365 1370 1375 1380 1385 1390 1395 1400 1405 1410 1415 1420 1425 1430 1435 1440 1445 1450 1455 1460 1465 1470 1475 1480 1485 1490 1495 1500 1505 1510 1515 1520 1525 1530 1535 1540 1545 1550 1555 1560 1565 1570 1575 1580 1585 1590 1595 1600 1605 1610 1615 1620 1625 1630 1635 1640 1645 1650 1655 1660 1665 1670 1675 1680 1685 1690 1695 1700 1705 1710 1715 1720 1725 1730 1735 1740 1745 1750 1755 1760 1765 1770 1775 1780 1785 1790 1795 1800 1805 1810 1815 1820 1825 1830 1835 1840 1845 1850 1855 1860 1865 1870 1875 1880 1885 1890 1895 1900 1905 1910 1915 1920 1925 1930 1935 1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020 2025 2030 2035 2040 2045 2050 2055 2060 2065 2070 2075 2080 2085 2090 2095 2100 2105 2110 2115 2120 2125 2130 2135 2140 2145 2150 2155 2160 2165 2170 2175 2180 2185 2190 2195 2200 2205 2210 2215 2220 2225 2230 2235 2240 2245 2250 2255 2260 2265 2270 2275 2280 2285 2290 2295 2300 2305 2310 2315 2320 2325 2330 2335 2340 2345 2350 2355 2360 2365 2370 2375 2380 2385 2390 2395 2400 2405 2410 2415 2420 2425 2430 2435 2440 2445 2450 2455 2460 2465 2470 2475 2480 2485 2490 2495 2500 2505 2510 2515 2520 2525 2530 2535 2540 2545 2550 2555 2560 2565 2570 2575 2580 2585 2590 2595 2600 2605 2610 2615 2620 2625 2630 2635 2640 2645 2650 2655 2660 2665 2670 2675 2680 2685 2690 2695 2700 2705 2710 2715 2720 2725 2730 2735 2740 2745 2750 2755 2760 2765 2770 2775 2780 2785 2790 2795 2800 2805 2810 2815 2820 2825 2830 2835 2840 2845 2850 2855 2860 2865 2870 2875 2880 2885 2890 2895 2900 2905 2910 2915 2920 2925 2930 2935 2940 2945 2950 2955 2960 2965 2970 2975 2980 2985 2990 2995 3000 3005 3010 3015 3020 3025 3030 3035 3040 3045 3050 3055 3060 3065 3070 3075 3080 3085 3090 3095 3100 3105 3110 3115 3120 3125 3130 3135 3140 3145 3150 3155 3160 3165 3170 3175 3180 3185 3190 3195 3200 3205 3210 3215 3220 3225 3230 3235 3240 3245 3250 3255 3260 3265 3270 3275 3280 3285 3290 3295 3300 3305 3310 3315 3320 3325 3330 3335 3340 3345 3350 3355 3360 3365 3370 3375 3380 3385 3390 3395 3400 3405 3410 3415 3420 3425 3430 3435 3440 3445 3450 3455 3460 3465 3470 3475 3480 3485 3490 3495 3500 3505 3510 3515 3520 3525 3530 3535 3540 3545 3550 3555 3560 3565 3570 3575 3580 3585 3590 3595 3600 3605 3610 3615 3620 3625 3630 3635 3640 3645 3650 3655 3660 3665 3670 3675 3680 3685 3690 3695 3700 3705 3710 3715 3720 3725 3730 3735 3740 3745 3750 3755 3760 3765 3770 3775 3780 3785 3790 3795 3800 3805 3810 3815 3820 3825 3830 3835 3840 3845 3850 3855 3860 3865 3870 3875 3880 3885 3890 3895 3900 3905 3910 3915 3920 3925 3930 3935 3940 3945 3950 3955 3960 3965 3970 3975 3980 3985 3990 3995 4000 4005 4010 4015 4020 4025 4030 4035 4040 4045 4050 4055 4060 4065 4070 4075 4080 4085 4090 4095 4100 4105 4110 4115 4120 4125 4130 4135 4140 4145 4150 4155 4160 4165 4170 4175 4180 4185 4190 4195 4200 4205 4210 4215 4220 4225 4230 4235 4240 4245 4250 4255 4260 4265 4270 4275 4280 4285 4290 4295 4300 4305 4310 4315 4320 4325 4330 4335 4340 4345 4350 4355 4360 4365 4370 4375 4380 4385 4390 4395 4400 4405 4410 4415 4420 4425 4430 4435 4440 4445 4450 4455 4460 4465 4470 4475 4480 4485 4490 4495 4500 4505 4510 4515 4520 4525 4530 4535 4540 4545 4550 4555 4560 4565 4570 4575 4580 4585 4590 4595 4600 4605 4610 4615 4620 4625 4630 4635 4640 4645 4650 4655 4660 4665 4670 4675 4680 4685 4690 4695 4700 4705 4710 4715 4720 4725 4730 4735 4740 4745 4750 4755 4760 4765 4770 4775 4780 4785 4790 4795 4800 4805 4810 4815 4820 4825 4830 4835 4840 4845 4850 4855 4860 4865 4870 4875 4880 4885 4890 4895 4900 4905 4910 4915 4920 4925 4930 4935 4940 4945 4950 4955 4960 4965 4970 4975 4980 4985 4990 4995 5000 5005 5010 5015 5020 5025 5030 5035 5040 5045 5050 5055 5060 5065 5070 5075 5080 5085 5090 5095 5100 5105 5110 5115 5120 5125 5130 5135 5140 5145 5150 5155 5160 5165 5170 5175 5180 5185 5190 5195 5200 5205 5210 5215 5220 5225 5230 5235 5240 5245 5250 5255 5260 5265 5270 5275 5280 5285 5290 5295 5300 5305 5310 5315 5320 5325 5330 5335 5340 5345 5350 5355 5360 5365 5370 5375 5380 5385 5390 5395 5400 5405 5410 5415 5420 5425 5430 5435 5440 5445 5450 5455 5460 5465 5470 5475 5480 5485 5490 5495 5500 5505 5510 5515 5520 5525 5530 5535 5540 5545 5550 5555 5560 5565 5570 5575 5580 5585 5590 5595 5600 5605 5610 5615 5620 5625 5630 5635 5640 5645 5650 5655 5660 5665 5670 5675 5680 5685 5690 5695 5700 5705 5710 5715 5720 5725 5730 5735 5740 5745 5750 5755 5760 5765 5770 5775 5780 5785 5790 5795 5800 5805 5810 5815 5820 5825 5830 5835 5840 5845 5850 5855 5860 5865 5870 5875 5880 5885 5890 5895 5900 5905 5910 5915 5920 5925 5930 5935 5940 5945 5950 5955 5960 5965 5970 5975 5980 5985 5990 5995 6000 6005 6010 6015 6020 6025 6030 6035 6040 6045 6050 6055 6060 6065 6070 6075 6080 6085 6090 6095 6100 6105 6110 6115 6120 6125 6130 6135 6140 6145 6150 6155 6160 6165 6170 6175 6180 6185 6190 6195 6200 6205 6210 6215 6220 6225 6230 6235 6240 6245 6250 6255 6260 6265 6270 6275 6280 6285 6290 6295 6300 6305 6310 6315 6320 6325 6330 6335 6340 6345 6350 6355 6360 6365 6370 6375 6380 6385 6390 6395 6400 6405 6410 6415 6420 6425 6430 6435 6440 6445 6450 6455 6460 6465 6470 6475 6480 6485 6490 6495 6500 6505 6510 6515 6520 6525 6530 6535 6540 6545 6550 6555 6560 6565 6570 6575 6580 6585 6590 6595 6600 6605 6610 6615 6620 6625 6630 6635 6640 6645 6650 6655 6660 6665 6670 6675 6680 6685 6690 6695 6700 6705 6710 6715 6720 6725 6730 6735 6740 6745 6750 6755 6760 6765 6770 6775 6780 6785 6790 6795 6800 6805 6810 6815 6820 6825 6830 6835 6840 6845 6850 6855 6860 6865 6870 6875 6880 6885 6890 6895 6900 6905 6910 6915 6920 6925 6930 6935 6940 6945 6950 6955 6960 6965 6970 6975 6980 6985 6990 6995 7000 7005 7010 7015 7020 7025 7030 7035 7040 7045 7050 7055 7060 7065 7070 7075 7080 7085 7090 7095 7100 7105 7110 7115 7120 7125 7130 7135 7140 7145 7150 7155 7160 7165 7170 7175 7180 7185 7190 7195 7200 7205 7210 7215 7220 7225 7230 7235 7240 7245 7250 7255 7260 7265 7270 7275 7280 7285 7290 7295 7300 7305 7310 7315 7320 7325 7330 7335 7340 7345 7350 7355 7360 7365 7370 7375 7380 7385 7390 7395 7400 7405 7410 7415 7420 7425 7430 7435 7440 7445 7450 7455 7460 7465 7470 7475 7480 7485 7490 7495 7500 7505 7510 7515 7520 7525 7530 7535 7540 7545 7550 7555 7560 7565 7570 7575 7580 7585 7590 7595 7600 7605 7610 7615 7620 7625 7630 7635 7640 7645 7650 7655 7660 7665 7670 7675 7680 7685 7690 7695 7700 7705 7710 7715 7720 7725 7730 7735 7740 7745 7750 7755 7760 7765 7770 7775 7780 7785 7790 7795 7800 7805 7810 7815 7820 7825 7830 7835 7840 7845 7850 7855 7860 7865 7870 7875 7880 7885 7890 7895 7900 7905 7910 7915 7920 7925 7930 7935 7940 7945 7950 7955 7960 7965 7970 7975 7980 7985 7990 7995 8000 8005 8010 8015 8020 8025 8030 8035 8040 8045 8050 8055 8060 8065 8070 8075 8080 8085 8090 8095 8100 8105 8110 8115 8120 8125 8130 8135 8140 8145 8150 8155 8160 8165 8170 8175 8180 8185 8190 8195 8200 8205 8210 8215 8220 8225 8230 8235 8240 8245 8250 8255 8260 8265 8270 8275 8280 8285 8290 8295 8300 8305 8310 8315 8320 8325 8330 8335 8340 8345 8350 8355 8360 8365 8370 8375 8380 8385 8390 8395 8400 8405 8410 8415 8420 8425 8430 8435 8440 8445 8450 8455 8460 8465 8470 8475 8480 8485 8490 8495 8500 8505 8510 8515 8520 8525 8530 8535 8540 8545 8550 8555 8560 8565 8570 8575 8580 8585 8590 8595 8600 8605 8610 8615 8620 8625 8630 8635 8640 8645 8650 8655 8660 8665 8670 8675 8680 8685 8690 8695 8700 8705 8710 8715 8720 8725 8730 8735 8740 8745 8750 8755 8760 8765 8770 8775 8780 8785 8790 8795 8800 8805 8810 8815 8820 8825 8830 8835 8840 8845 8850 8855 8860 8865 8870 8875 8880 8885 8890 8895 8900 8905 8910 8915 8920 8925 8930 8935 8940 8945 8950 8955 8960 8965 8970 8975 8980 8985 8990 8995 9000 9005 9010 9015 9020 9025 9030 9035 9040 9045 9050 9055 9060 9065 9070 9075 9080 9085 9090 9095 9100 9105 9110 9115 9120 9125 9130 9135 9140 9145 9150 9155 9160 9165 9170 9175 9180 9185 9190 9195 9200 9205 9210 9215 9220 9225 9230 9235 9240 9245 9250 9255 9260 9265 9270 9275 9280 9285 9290 9295 9300 9305 9310 9315 9320 9325 9330 9335 9340 9345 9350 9355 9360 9365 9370 9375 9380 9385 9390 9395 9400 9405 9410 9415 9420 9425 9430 9435 9440 9445 9450 9455 9460 9465 9470 9475 9480 9485 9490 9495 9500 9505 9510 9515 9520 9525 9530 9535 9540 9545 9550 9555 9560 9565 9570 9575 9580 9585 9590 9595 9600 9605 9610 9615 9620 9625 9630 9635 9640 9645 9650 9655 9660 9665 9670 9675 9680 9685 9690 9695 9700 9705 9710 9715 9720 9725 9730 9735 9740

memory 250, to scan and/or store only specific portions of hard copy document 100. After the interactive prompts required to obtain information for a desired application program, the unused information stored in scanner memory 220 or 250 can be erased. Further, scanning of a second identical document can be limited to only those portions of the document which contain needed information.

More specifically, in FIG. 2, the lines 10 drawn around certain portions of the document represent the areas which the user has previously identified as the portions of a document to be extracted by the scanner 210 and stored in scanner memory 220 and/or main memory 250. Since the logo 20 and the message 30 have not been identified as an area to be scanned and stored, these areas are not scanned and stored in subsequent documents. Since the user has previously associated each of the areas 10 with a specific subfile of information, e.g., the account number, the scanned information is stored in memory locations corresponding to that subfile.

Data Processing

FIGS. 4-10 illustrate the flow of data in the FIG. 1 preferred embodiment. FIG. 4 illustrates the overall data flow for the FIG. 1 preferred embodiment. The preferred embodiment includes an input process module 1.0, an information processing module 2.0, and an output processing module 3.0. The information processing module 2.0 is equipped to receive instructions from and transmit information to a user. The information processing module 2.0 can also transmit to and receive information from a remote external device through communication interface 4.0. Input process module 1.0 and output processing module 3.0 can also access communication interface 4.0. A module is implemented in hardware, software, or a combination of hardware and software. The specific implementation for a particular business application depends upon a variety of factors, for example, the relative costs of hardware and software implemented systems, the frequency with which a user will want to expand or modify the system, and the like.

FIG. 5 is a more detailed diagram of the input process module 1.0 of FIG. 4. The input process module 1.0 includes a character input module 1.1, an image input module 1.2, and, in the preferred embodiment, a character recognition device 1.3. The character input module inputs textual information, such as alphanumeric characters, from an input device such as keyboard 234. The image input module 1.2 inputs image information, for example, a digitized image of the actual appearance of hard copy document 100. Textual information can include textual input from an input device such as keyboard 234 and textual information extracted from the document by character recognition device 1.3. Both types of information comprise an input document which is transmitted to information processing module 2.0. In the FIG. 1 preferred embodiment, the processing performed by input process module 1.0 occurs in scanner memory 220, computer 230, and main memory 250.

FIG. 6 illustrates information processing data flow for the FIG. 1 preferred embodiment, that is, FIG. 6 illustrates data flow in the information processing module 2.0.

The information processing module 2.0 includes a maintain library module 2.1, to be described in further detail below in conjunction with FIG. 7, a maintain definitions module 2.2, to be described in further detail below in conjunction with FIG. 8, and a process document module 2.3 to be described in further detail below in conjunction with FIG. 9.

The information processing module 2.0 is the module which coordinates and drives the entire system. In the preferred embodiment, the information processing module 2.0 is implemented primarily by computer 230.

FIG. 7 illustrates information processing data flow in the maintain library module 2.1. The maintain library module 2.1 maintains a library of image information, for example, a digitized image representing the actual appearance of the hard copy document, and textual information of the hard copy documents for reference during processing. This library can be incorporated within scanner memory 220, main memory 250, or another independent memory, for example, a RAM disk. The maintain library module 2.1 includes a store document module 2.1.1, a correct errors module 2.1.2, a retrieve document module 2.1.3, and a document file 2.1.4. These modules operate collectively to store, retrieve, and correct document information.

The store document module 2.1.1, prior to routing the document to the document file 2.1.4, may provide information on recognition errors which may have occurred while inputting the document. For example, the store document module 2.1.1 identifies that a character contained on hard copy document 100 was not recognized. The store document module 2.1.1 also optionally causes a copy of the document and its parsing to be displayed on the display 232 for confirmation by the user. The user may utilize this opportunity to identify any errors in the displayed document and, in conjunction with the correct errors module 2.1.2, to revise the document's parsing, if necessary, prior to storage of the document in memory. The module 2.1.1 also provides a facility for the user to name a particular hard copy document for cataloging, storage, and retrieval purposes. After the document is named, the store document module 2.1.1 stores copies of the document in the document file 2.1.4.

The correct errors module 2.1.2 processes instructions from the user to correct errors identified by the store document module 2.1.1 and errors that have been spotted by the user during the confirmation process.

The retrieve document module 2.1.3 permits the user to retrieve a copy of a document previously stored in the document file 2.1.4. As described above, long-term storage is provided by main memory 250, if necessary.

FIG. 8 illustrates a more detailed information processing data flow diagram for the maintain definitions module 2.2 of FIG. 6. The maintain definitions module 2.2 allows the user to define system and document parameters and maintains the definitions of these system and document parameters. The maintain definitions module 2.2 includes a define template module 2.2.1 which allows the user to specify the location of information on the document. This information provided by the user defines a template which is used to extract information off the document and to associate the extracted information with a particular variable or subfile. These templates are illustrated by boxes 10 in the FIG. 2 example of a hard copy document. The maintain definitions module 2.2 can also access templates previously defined by the user and stored in main memory 250. Templates can also be provided as part of software packages developed by program developers.

The maintain definitions module 2.2 also includes a define relationships module 2.2.2. The define relationships module 2.2.2 allows the user to define data relationships, or logical relationships, between pieces of information extracted from the hard copy document. These pieces of information are then used to generate an input file for a selected computer application unit. The user defines these relationships by

content instructions. Alternatively, content instructions to define relationships can be provided by application software. If the user provides these content instructions, the content instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. Examples of content instructions, data, and logical relationships will be described in further detail in conjunction with Appendices 11 and 12A, 12B, and 12C.

The maintain definitions module 2.2 also includes a define format module 2.2.3. The define format module 2.2.3 allows the user to define transmission formats for an input file which is then transmitted to a selected computer application unit. Selection of the transmission format of the input file is accomplished by the user through use of transmission format instructions. Alternatively, the applications software itself can generate its own transmission format instructions. When the user must specify transmission format instructions, the transmission format instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. A further description of various transmission formats will be provided below in conjunction with Appendices 12A, 12B, 12C, 13A, 13B, and 13C.

A select definitions module 2.2.4 is also included in the maintain definitions module. The select definitions module 2.2.4 allows the user to store and select a set of definitions to be used for processing the document. The definitions identify pieces of information on the document by, for example, absolute location, variable location, or relative location, or by proximity to key words and/or symbols. These definitions are described in further detail below by way of an illustrative example.

FIG. 9 illustrates a more detailed information processing data flow diagram for the process document module 2.3. The process document module 2.3 processes the document after the document has been stored in the system. The process document module 2.3 gathers the appropriate information which has been stored, and creates input file(s) 2.3.3 for the selected application unit. The process document module 2.3 then transmits the input file(s) via bus 240 and/or communication interface 4.0 to an application unit 270, an output device such as printer 260, or to main memory 250.

The process document module 2.3 includes an extract data module 2.3.1. This module extracts data off of the document in accordance with the user's instructions, for example, the user-defined template, or through the interactive mode.

The process document module 2.3 also includes a preapplication process module 2.3.2 which gathers and associates information extracted from the document in accordance with content instructions. This module prompts the user for any additional information required to satisfy the relationships defined by the content instructions. The preapplication process module 2.3.2 also places the selected information into the transmission format defined by the transmission format instructions.

The preapplication process module 2.3.2 also generates the input file 2.3.3 for the selected application in accordance with the appropriate instructions. The input file 2.3.3 is then transmitted to bus 240 and/or communication interface 4.0 for transmission to a particular application unit 270.

FIG. 10 illustrates a detailed output data flow diagram for output module 3.0. Output module 3.0 outputs a textual and/or image copy of the document. In the FIG. 1 preferred embodiment, output module 3.0 is implemented by printer 260, associated software, and associated interface circuitry.

Operation

Examples of operation of a preferred embodiment will now be described.

5 The user enters the system by providing instructions to the information processing module 2.0. The user then instructs the information processing module 2.0 to conduct maintain library processing, maintain definitions processing, or process document processing.

10 If the user selects maintain library processing, the user then provides instructions to maintain or modify the document library through the maintain library module 2.1. For example, the user can direct the inputting and storage of a hard copy document 100 or can retrieve and output a document. The user requests inputting of a document through the store document module 2.1.1. The system then prompts the user to specify a storage location for the inputted document. The document is then read in by the input process module 1.0. A textual copy and/or an image copy are stored into the document file 2.1.4. Errors which have occurred during inputting are identified and corrected by the correct errors module 2.1.2 and the user. The corrections are reflected in the document information stored in document file 2.1.4.

15 The retrieve document module 2.1.3 is used to retrieve and output a document. The system prompts the user to specify the storage location of a document and the type of document copy, for example, a textual or an image copy, to be outputted. The document is then outputted by the output process module 3.0.

20 30 If the user initially selected maintain definitions processing, the user would instruct the system to maintain and/or modify parameter definitions through the maintain definitions module 2.2. For example, the user can define and maintain a document template for extracting selected portions of information off of the hard copy document. The user can use the template to extract selected portions of information off of the hard copy document when the document is originally inputted, or alternatively, the user can use the template to identify selected portions of information for extraction off of an image copy of the document. In creating the template, the user identifies pieces of information on the document to be extracted and assigns a variable name, or subfile, to each piece of data.

25 35 40 The location of data to be extracted can be defined in a number of ways other than by use of a template. For example, the user can designate the absolute location of information on the document with respect to a grid overlaid on the document, e.g., always on line 3, starting in column 1. The user can also identify information by specifying the relative location of information to be extracted, e.g., always two lines below the piece of data named "salutation" starting in column 3. The user can also specify the location of information to be extracted by variable location specification. For example, if the hard copy document is a letter, the module would conduct a key word search for the term "Dear Sir:". Wherever this term "Dear Sir:" is located, this piece of data would be associated with the variable specified by the user, for example, the variable "salutation." In addition, a defined set of conventional symbols can be used to signify certain recurring data items for the convenience of users of the instant invention. For example, a "@" symbol can be used to delineate the vendor name as follows: "@XYZ Corporation@". Other examples of the use of symbols to delineate information will be described with reference to Appendix 4.

45 50 55 60 65 The maintain definition module 2.2 is also used to maintain data relationships in accordance with content instruc-

tions and to maintain input file formats in accordance with transmission format instructions. Relationships are defined and maintained between pieces of data, specified by, for example, the names of variables, through the define relationships module 2.2.2. The names of pieces of data on the document are retrieved by, for example, the define template module 2.2.1, and are passed to the define relationships module 2.2.2. The user may then provide any additional pieces of data needed to generate an input file for a particular application program or unit, such as an input file line number. The user, the applications software, and/or instructions previously stored in memory then establishes the contents of the input file by defining relationships between pieces of data using content instructions. Specific examples of content instructions will be discussed below in conjunction with Appendices 11, 12A, 12B, 12C, 13A, 13B, and 13C.

The user and/or the applications software defines and maintains the transmission format of the input file to be used by a particular application program or unit through the define format module 2.2.3 in accordance with transmission format instructions. This is accomplished by defining the parameters to be used by the preapplication process module 2.3.2 in generating an input file. Parameters which would typically be required to generate an input file would include the character type, e.g., text or pixel; delimiters used between pieces of data, e.g., a slash or a semicolon; end of line characters, e.g., a carriage return or a line feed; and end of file characters. Examples of transmission formats will be described in further detail below in conjunction with Appendices 11, 12A, 12B, 12C, 13A, 13B, and 13C.

If the user initially selected process document processing, the interface will then proceed to process the document through use of the process document module 2.3. For example, the user can extract specific portions of data from an image copy of a document, can generate an input file for transmission to an application program, or can directly process information interactively with an application program.

If the user desires to extract specific portions of data from an image copy of a hard copy document which has already been stored in memory, the user uses the extract data module 2.3.1 to identify a document to be processed. The document is then retrieved by the retrieve document module 2.1.3 and passed to the extract data module 2.3.1. The user can also select parameter definitions through the select definitions module 2.2.4.

The selected document template or parameter definition is passed to the extract data module 2.3.1. The extract data module 2.3.1 extracts pieces of data from the image copy of the document, as defined by the document template definition or the parameter definitions or both. This document data is then passed to preapplication process module 2.3.2.

The interface generates input file(s) 2.3.3 by use of the preapplication process module 2.3.2. The selected data relationship definition, as defined by the content instructions, and the selected record format definitions, as defined by the transmission format instructions, are passed to the preapplication process module 2.3.2. The preapplication process module 2.3.2 assembles the input file in accordance with the content instructions. The preapplication process module 2.3.2 also prompts the user for any additional pieces of data which need to be provided by the user. The input file is converted to the desired transmission format in accordance with the transmission format instructions. This physically formatted data is then stored in the input file 2.3.3.

The user can also use an application program to process information by loading the particular application program into the computer 230 rather than by sending the input file to a remote application unit 270.

5 An illustrative example of the processing described above will now be described.

The user inputs instructions via keyboard 234 or another input device which indicate that the user desires to input and store a document. The computer 230 then prompts the user for the name of the document. In this example, the user desires to input the document of FIG. 2 and therefore names the document "XYZ Corp. Bill Dec. 1, 1986." The computer then prompts the user to feed the hard copy document 100 into the scanner 210. The image of the hard copy document is displayed on display 232. The computer then prompts the user to identify the account number on the document. By use of the mouse 236 or other input device to position a cursor on the display, the user indicates the location of the account number. The account number is then read-in to a subfile named "Account Number." This process proceeds until all of the desired information has been read-in and stored.

In this particular example, no errors were encountered while inputting the document. The user then directs that the document be stored for future reference in a document file.

25 Some time later, the user desires to retrieve and output the document and to generate input files based on information from the document. The computer 230 prompts the user for the name of the document and the type of output. The user responds with "XYZ Corp. Bill Dec. 1, 1986" for a printed textual copy. The document is then retrieved from the document file and passed to the printer 260 for printing.

In order to generate an input file for a specific application program, the user selects the option to define a document template for use when each month's XYZ Corporation bill arrives. Accordingly, the user instructs the system to display a copy of an XYZ Corporation bill on the display 232. The user then identifies pieces of data by absolute locations. That is, the user assigns specific names to information located at 30 specific portions of the document. In this example, the user would input the following information:

40 Vendor-text, line 1, one line, column 1, 80 characters;
Account number-numeric, line 6, one line, column 25, 9 characters;

Statement date-date, line 9, one line, column 25, 8 characters;

Payment date-date, line 11, one line, column 25, 8 characters;

Previous balance-currency, line 7, one line, column 75, 9 characters;

New charges-currency, line 8, one line, column 75, 9 characters;

Other debits-currency, line 10, one line, column 75, 9 characters;

Finance charges-currency, line 12, one line, column 75, 9 characters;

Payments-currency, line 13, one line, column 75, 9 characters;

Other credits-currency, line 14, one line, column 75, 9 characters;

New balance-currency, line 15, one line, column 75, 9 characters.

The user also identifies data with variable locations. In this particular example, a variable location is specified as follows:

13

Heading 2-line, value="Mail To:"

The identification of Heading 2 as line information means that the system will search for occurrences of the character string "Mail To:" and assign the line number which contains this character string to Heading 2.

The user also identifies data by relative locations. In this example, the user identifies the following relative location:

Mail To-text, Heading 2+1, 3 lines, column 60, 25 characters per line.

The instructions above instruct the system to assign the textual information beginning on one line after Heading 2 and continuing for 3 lines, in column 60, to the Mail To subfile.

As an alternative to inputting the actual line, column, and character numbers, the user can identify desired portions of the document by blocking, or highlighting, the desired portions using the mouse or other input device. In this case, the computer converts the highlighted portions into corresponding line, column, and character numbers.

Appendix 11 lists data corresponding to the hard copy document of FIG. 2 and the associated variable or subfile names.

Next, the user desires to define data relationships in accordance with content instructions. Examples of the type of contents which can be specified by a user are illustrated in Appendices 12A, 12B, and 12C.

In this particular example, three separate departments of ABC Corporation require information from the XYZ Corporation bill. The first department requires vendor, account number, statement date, payment date, previous balance, new charges, debits, finance charges, payments, and new balance information. The second and third departments require mail to information and previous balance information. Each of these departments have their own application program which utilizes this information.

The user employs content instructions to designate how pieces of information, which have been extracted off of hard copy document 100, are directed to particular departments, that is, particular application programs. Appendix 12A illustrates the contents of the information to be transmitted to the first department. FIG. 12B illustrates the information to be transmitted to the second department. Appendix 12C illustrates the information to be transmitted to the third department. The content instructions, therefore, parse the information shown in FIG. 11 to various application programs, as shown by Appendices 12A, 12B, and 12C. Content instructions can also be used to identify additional pieces of data which are required for the input files of the particular application programs. In this particular example, the specific application programs from the three departments all require numeric record number information, numeric horizontal position information, numeric vertical position information, and date received information. The horizontal and vertical position information is used by the application program to specify the location of the received information on a spreadsheet application program, in this example. The user may know in advance the content format required by each application program, that is, in this example, the location and type of information specified on the spreadsheet. The user may also employ the split display mode described with reference to FIG. 3B to generate content format instructions.

Using the content instructions, the user establishes the following contents for the input file corresponding to Appendices 12A:

Record number, horizontal position, vertical position, vendor;

Record number, horizontal position, vertical position, account number;

14

Record number, horizontal position, vertical position, statement date;

Record number, horizontal position, vertical position, date received;

Record number, horizontal position, vertical position, payment date;

Record number, horizontal position, vertical position, previous balance;

Record number, horizontal position, vertical position, new charges;

Record number, horizontal position, vertical position, finance charges;

Record number, horizontal position, vertical position, payments;

Record number, horizontal position, vertical position, new balance.

Next, transmission format instructions are employed to define the transmission format of the input file for a specific application program or unit. Appendix illustrates the transmission input file corresponding to Appendix 12A. Appendix FIG. 13B illustrates the transmission input file corresponding to Appendix 12B. Appendix 13C illustrates the transmission input file corresponding to Appendix 12C. A comparison of Appendices 12B and 12C reveals that Appendices 12B and 12C have the same contents. However, the information illustrated in FIG. 12B is being sent to a different application program than the information in Appendix 12C. These application programs require different transmission input formats, as illustrated in Appendices 13B and 13C. More specifically, the application program that receives the input file illustrated in Appendix 13B uses the greater than sign as a delimiter whereas the application program which receives the transmission input file shown in Appendix 13C uses a back-slash as the delimiter.

After the contents and the transmission format for the input file have been defined, and any additional information has been inputted, the input file is assembled and transmitted to the particular application program.

Appendix 14 illustrates another example of a hard copy document containing information to be processed by the instant invention. The hard copy document illustrated in Appendix 14 is first scanned and information from the hard copy document is stored into a memory. The interface 200 then identifies portions of the hard copy document corresponding to various variables by recognizing a defined set of symbols.

In the Appendix 14 example, triangles delineate the mailing address, circles delineate the statement date, and squares delineate the new charges. Information from these portions of the hard copy document is stored in the corresponding memory locations or subfiles for each variable. The same set of symbols can be used to identify the same information from one document to the next. Thus, even if the physical formats of documents are not fixed from one document to the next, a diversity of hard copy documents can be processed without manually inputting data by recognition of the defined symbols.

Examples of readily available application programs are Quicken and Lotus 1.2.3 both of which are widely utilized in the business community. Quicken, for example is an easy-to-utilize program for writing checks and preparing business records. Payee, amount and address information may readily be transmitted from scanner memory 220 and/or main memory 250 to the Quicken application program for check writing functions and ledger keeping purposes. Lotus

is a well known spreadsheet program which may process data input into specified cells once this data is placed in conventional Lotus format.

Thus, the instant invention provides an integrated and comprehensive system for handling information from a hard copy document, thus permitting a paperless office. In addition, the invention permits data, extracted off of a hard copy document, to be easily manipulated into various logical and transmission formats required by a particular application unit. The invention also provides a low cost system for inputting information from a wide variety of hard copy documents into a memory.

The foregoing description has been set forth merely to illustrate preferred embodiments of the invention and is not intended to be limiting. Modifications are possible without departing from the scope of the invention.

For example, letters, checks, forms, pictures, reports, music scores, film, and other types of hard copy documents can be processed by the invention for accounts payable/receivable accounting, inventory control, record keeping, budgeting, data base management, music transcription, forms processing, computerized art, survey and questionnaire processing, statistical data analysis, correspondence processing and other applications.

Other automated digitizing units can be used in addition to or as an alternative to use of the scanner 210 as an input unit. Any electrical, magnetic, or optical device which

extracts information off of a hard copy document, thereby eliminating the need to manually input significant amounts of information from the hard copy document is suitable for use as an automated digitizing unit. In addition, information 5 can be input by user responses and digital and analog signals generated from various devices, and from computer files from other computer systems. Suitable hardware for inputting data includes a keyboard, a light pen, a mouse, a touch screen, a laser scanner, a microphone, a tablet, a disk drive, a magnetic tape drive, and a modem.

The interface 200 can also output information in forms other than a hard copy of textual or image information. For example, the interface 200 can output system responses, computer files, and digital and analog signals for transmission to other computer systems or to control systems. Suitable hardware for outputting information includes a disk drive, a magnetic tape drive, a cathode ray tube, a plasma screen, a printer, a plotter, a film developer, an amplifier, and a modem.

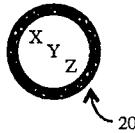
Since modifications of the described embodiments incorporating the spirit and substance of the invention may occur to persons skilled in the art, the scope of the invention should be limited solely with respect to the appended claims and equivalents.

Appendix
(No. 11)

	Variable Name	Value
	Vendor	XYZ Corporation
	Heading 2	2
	Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
	Account Number	123456789
	Statement Date	12/01/86
	Payment Date	1/01/87
	Previous Balance	\$1234.56
	New Charges	\$789.01
	Debits	
	Finance Charges	\$2.34
	Payments	\$1000.00
	Other Credits	
	New Balance	\$1025.91
No. 12A	Vendor	XYZ-Corporation
	Account Number	123456789
	Statement Date	12/01/86
	Payment Date	1/01/87
	Previous Balance	\$1234.56
	New Charges	\$789.01
	Debits	
	Finance Charges	\$2.34
	Payments	\$1000.00
	Other Credits	
	New Balance	\$1025.91
No. 12B	Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
	Previous Balance	\$1234.56
	Mail To	XYZ corporation PO Box 567 Anywhere, NY 63130
No. 12C	Previous Balance	\$1234.56
No. 13A	Previous Balance	\$1234.56 >1>1>1>"XYZ Corporation" >2>2>2>+123456789> >3>2>1>D12/01/86> >4>2>11>D12/15/86> >5>2>21>D01/01/87> >6>10>25>\$1234.56> >7>11>25>\$789.01> >8>13>25>\$2.34> >9>14>25>\$1000.00> >10>16>25>\$1025.91> >1>1>1>\$1234.56>
No. 13B		

No. 13C

>2>2>1>“XYZ Corporation”
 >3>3>1>“PO Box 567”
 >4>4>1>“Anywhere, NY 63130”
 /1/1/1/\$1234.56/
 /2/2/1/*XYZ Corporation*
 /3/3/1/*PO Box 567*
 /4/4/1/*Anywhere, NY 63130*
 Appendix
 No. 14



XYZ Corporation

20

Customer

ABC Corporation
 123 Sixteenth Street
 Hometown, NJ 88981
 Account Number 123456789

Statement Date:	12/81/86	Previous Balance	\$1234.56
Payment Date:	1/81/87	New Charges	\$789.01
		Other Debits	
		Finance Charges (10%)	\$2.34
		Payments	\$1888.88
		Other Credits	
		New Balance	\$1825.91

WINTER SALE IN EFFECT 30
THROUGHOUT JANUARY!

Mail To:
 XYZ Corporation
 PO Box 567
 Anywhere, NY 63130

30

What is claimed is:

1. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

(a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;

35

(b) a display to display information;
 (c) an input device to receive user instructions from a user; and
 (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in at least one of:

40

(1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;

45

(2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and

(3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program;

wherein the display projects an image showing at least a portion of said document information and at least one field of information required by said at least one application program during the interactive mode.

2. A multimode information processing system as set forth in claim 1, wherein said system further comprises:

a memory for storing a defined set of symbols which designate fields of information required by said at least one application program; and

a symbol detection means for detecting the presence of a particular one of said defined set of symbols within said document information and for extracting a field of information required by said at least one application program based on said detecting.

3. A multimode information processing system as set forth in claim 1 wherein said system is operative in said interactive mode for locating said at least one field of information required by said at least one application program by blocking portions of said at least a portion of said document information displayed on the display using the input device.

4. A multimode information processing system as set forth in claim 1 wherein said system is operative in said interactive template creation mode for creating a template by blocking portions of said at least a portion of said document information displayed on the display using the input device.

5. A multimode information processing system as set forth in claim 1, further comprising:

a memory for storing digitized image information representative of the actual appearance of at least a portion of said hard copy document.

65

19

6. A multimode information processing system as set forth in claim 1, further comprising:

a memory for storing at least a portion of said textual data.

7. A multimode information processing system as set forth in claim 1, further comprising:

a memory for storing digitized image information representative of the actual appearance of at least a portion of said hard copy document; and

a memory for storing at least a portion of said textual data.

8. A multimode information processing system as set forth in claim 1, wherein said character recognition capabilities are included in the automated digitizing unit.

9. A multimode information processing system as set forth in claim 1 wherein in said interactive template creation mode, said template identifies a plurality of locations within said document information of a plurality of fields of information required by an application program.

10. A multimode information processing system as set forth in claim 1 wherein said at least one application program is a plurality of application programs.

11. A multimode information processing system as set forth in claim 1 wherein in said interactive template creation mode, said template identifies a plurality of locations within said document information of a plurality of fields of information required by a plurality of application programs.

12. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

(a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;

(b) a display to display information;

(c) an input device to receive user instructions from a user; and

(d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in at least one of:

(1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;

(2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and

(3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of

20

information required by the application program and to transfer said fields of information to said at least one application program; and

wherein the display projects an image showing at least a portion of said document information and at least one field of information required by said at least one application program during said interactive template creation mode.

13. A multimode information processing system as set forth in claim 12 wherein said system is operative in said interactive template creation mode for creating a template by blocking portions of said at least a portion of said document information displayed on the display using the input device.

14. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

(a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;

(b) a display to display information;

(c) an input device to receive user instructions from a user; and

(d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in at least one of:

(1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;

(2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and

(3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein, in said interactive mode, said located information is located for and transferred to at least two fields.

15. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

(a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;

(b) a display to display information;

(c) an input device to receive user instructions from a user; and

21

(d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by said at least one application program to said at least one application program, and to operate in at least two of:

- (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;
- (2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and
- (3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein the display projects an image showing at least a portion of said document information and at least one field of information required by said at least one application program during the interactive mode.

16. A multimode information processing system as set forth in claim 15 wherein said system further comprises:

a memory for storing a defined set of symbols which designate fields of information required by said at least one application program; and

a symbol detection means for detecting the presence of a particular one of said defined set of symbols within said document information and for extracting a field of information required by said at least one application program based on said detecting.

17. A multimode information processing system as set forth in claim 15 wherein said computer is operative in said interactive mode for locating said at least one field of information required by said at least one application program by blocking portions of said at least a portion of said document information displayed on the display using the input device.

18. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;
- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device;

22

said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by said at least one application program to said at least one application program, and to operate in at least two of:

- (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;
- (2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and
- (3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein said computer is operative in said interactive mode for locating said at least one field of information required by said at least one application program by blocking portions of said at least a portion of said document information displayed on the display using the input device.

19. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;
- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by said at least one application program to said at least one application program, and to operate in at least two of:

- (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;

(2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and

(3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein the display projects an image showing at least a portion of said document information and at least one field of information required by said at least one application program during said interactive template creation mode.

20. A multimode information processing system as set forth in claim **19** wherein said computer is operative in said interactive template creation mode for creating a template by blocking portions of said at least a portion of said document information displayed on the display using the input device.

21. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;
- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by said at least one application program to said at least one application program, and to operate in at least two of:
 - (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;
 - (2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and
 - (3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and

to transfer said fields of information to said at least one application program; and

wherein said computer is operative in said interactive template creation mode for creating a template by blocking portions of said at least a portion of said document information displayed on the display using the input device.

22. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;
- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by said at least one application program to said at least one application program, and to operate in at least two of:

(1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;

(2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and

(3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

further comprising a memory for storing digitized image information representative of the actual appearance of at least a portion of said hard copy document.

23. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;
- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric charac-

ters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by said at least one application program to said at least one application program, and to operate in at least two of:

- (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;
- (2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and
- (3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

further comprising a memory for storing at least a portion of said textual data.

24. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;
- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by said at least one application program to said at least one application program, and to operate in at least two of:

- (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;
- (2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and

5

10

15

20

25

30

40

45

55

60

65

and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and

- (3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

further comprising:

a memory for storing digitized image information representative of the actual appearance of at least a portion of said hard copy document; and

a memory for storing at least a portion of said textual data.

25. A multimode information processing system as set forth in claim 24 wherein said character recognition capabilities are included in the automated digitizing unit.

26. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;
- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by said at least one application program to said at least one application program, and to operate in at least two of:

(1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;

(2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and

(3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein, in said interactive mode, said located information is located for and transferred to at least two fields.

27. A multimode information processing system as set forth in claim 26 wherein in said interactive template

creation mode, said template identifies a plurality of locations within said document information of a plurality of fields of information required by an application program.

28. A multimode information processing system as set forth in claim 26 wherein said at least one application program is a plurality of application programs. 5

29. A multimode information processing system as set forth in claim 26 wherein in said interactive template creation mode, said template identifies a plurality of locations within said document information of a plurality of 10 fields of information required by a plurality of application programs.

30. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising: 15

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;
- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and 20
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for 25 recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in each of:
- (1) an interactive mode wherein at least a portion of 30 said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;
- (2) an interactive template creation mode wherein an 35 image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and
- (3) an automatic mode to match at least a portion of 40 said document information with a template stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and 45

wherein the display projects an image showing at least a portion of said document information and at least one field of information required by said at least one application program during the interactive mode.

31. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;
- (b) a display to display information;

(c) an input device to receive user instructions from a user; and

(d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in each of:

- (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;
- (2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and
- (3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein the display projects an image showing at least a portion of said document information and at least one field of information required by said at least one application program during the interactive mode.

32. A multimode information processing system as set forth in claim 30 or claim 31, wherein said system further comprises:

- a memory for storing a defined set of symbols which designate fields of information required by said at least one application program; and
- a symbol detection means for detecting the presence of a particular one of said defined set of symbols within said document information and for extracting a field of information required by said at least one application program based on said detecting.

33. A multimode information processing system as set forth in claim 30 or claim 31, further comprising:

- a memory for storing digitized image information representative of the actual appearance of at least a portion of said hard copy document.

34. A multimode information processing system as set forth in claim 30 or claim 31, further comprising:

- a memory for storing at least a portion of said textual data.

35. A multimode information processing system as set forth in claim 30 or claim 31, wherein said character recognition capabilities are included in the automated digitizing unit.

36. A multimode information processing system as set forth in claim 30 or claim 31, wherein in said interactive

template creation mode, said template identifies a plurality of locations within said document information of a plurality of fields of information required by an application program.

37. A multimode information processing system as set forth in claim 30 or claim 31, wherein said at least one application program is a plurality of application programs. 5

38. A multimode information processing system as set forth in claim 30 or claim 31, wherein in said interactive template creation mode, said template identifies a plurality of locations within said document information of a plurality 10 of fields of information required by a plurality of application programs.

39. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising: 15

(a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;

(b) a display to display information;

(c) an input device to receive user instructions from a user; and

(d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in each of: 25

(1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program; 35

(2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and 40

(3) an automatic mode to match at least a portion of said document information with a template stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and 45

wherein said computer is operative in said interactive mode for locating said at least one field of information required by said at least one application program by blocking portions of said at least a portion of said document information displayed on the display using the input device. 60

40. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

(a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;

(b) a display to display information;

(c) an input device to receive user instructions from a user; and

(d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in each of:

(1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;

(2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and

(3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein said computer is operative in said interactive mode for locating said at least one field of information required by said at least one application program by blocking portions of said at least a portion of said document information displayed on the display using the input device.

41. A multimode information processing system as set forth in claim 30 or claim 31, further comprising:

a memory for storing digitized image information representative of the actual appearance of at least a portion of said hard copy document; and

a memory for storing at least a portion of said textual data.

42. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

(a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;

(b) a display to display information;

(c) an input device to receive user instructions from a user; and

(d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information.

31

and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in each of:

5 (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;

(2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and

(3) an automatic mode to match at least a portion of said document information with a template stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein the display projects an image showing at least a portion of said document information and at least one field of information required by said at least one application program during said interactive template creation mode.

43. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

(a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;

(b) a display to display information;

(c) an input device to receive user instructions from a user; and

(d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in each of:

(1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;

(2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program;

55 60 65

32

least one field of information required by said at least one application program; and

(3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

10 wherein the display projects an image showing at least a portion of said document information and at least one field of information required by said at least one application program during said interactive template creation mode.

44. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

(a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;

(b) a display to display information;

(c) an input device to receive user instructions from a user; and

(d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in each of:

(1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;

(2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and

(3) an automatic mode to match at least a portion of said document information with a template stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein said system is operative in said interactive template creation mode for creating a template by blocking portions of said at least a portion of said document information displayed on the display using the input device.

45. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

(a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;

- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in each of:
 - (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;
 - (2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and
 - (3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein said system is operative in said interactive template creation mode for creating a template by blocking portions of said at least a portion of said document information displayed on the display using the input device.

46. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;
- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in each of:
 - (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from

5

10

15

20

25

30

35

40

45

50

55

60

65

the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;

- (2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and
- (3) an automatic mode to match at least a portion of said document information with a template stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein, in said interactive mode, said located information is located for and transferred to at least two fields.

47. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;
- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in each of:
 - (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;
 - (2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and
 - (3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein, in said interactive mode, said located information is located for and transferred to at least two fields.

48. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document; 5
- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and 10
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data. said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at 15 least one application program, and to operate in at least one of:
 - (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at 20 least one application program; 25
 - (2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and 30
 - (3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and 35

wherein the display projects an image showing at least a portion of said document information and at least one field of information required by said at least one application program during said automatic mode.

49. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document; 55
- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data. said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of 60

said document information as input data to be processed by said at least one application program to said at least one application program, and to operate in at least two of:

- (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;
- (2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and
- (3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein the display projects an image showing at least a portion of said document information and at least one field of information required by said at least one application program during at least said automatic mode.

50. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

- (a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;
- (b) a display to display information;
- (c) an input device to receive user instructions from a user; and
- (d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data. said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in each of:
 - (1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;
 - (2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at 65

least one field of information required by said at least one application program; and

(3) an automatic mode to match at least a portion of said document information with a template stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein the display projects an image showing at least a portion of said document information and at least one field of information required by said at least one application program during at least said automatic mode.

51. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

(a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;

(b) a display to display information;

(c) an input device to receive user instructions from a user; and

(d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in each of:

(1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;

(2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and

(3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein the display projects an image showing at least a portion of said document information and at least one field of information required by said at least one application program during said automatic mode.

52. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

(a) an automated digitizing unit providing digitized image information representative of an image of at least a portion of said document;

(b) a display to display information;

(c) an input device to receive user instructions from a user; and

(d) a computer, operable with at least one of the automated digitizing unit, the display, and the input device; said system having character recognition capabilities for recognizing at least a portion of alpha/numeric characters appearing on said document for providing textual data, said digitized image information and said textual data collectively comprising document information, and said system configured to route at least a portion of said document information as input data to be processed by at least one application program to said at least one application program, and to operate in at least one of:

(1) an interactive mode wherein at least a portion of said document information is displayed on the display and the input device conveys instructions from the user for locating information required by the application program within said document information and for transferring said located information to at least one field of information required by said at least one application program;

(2) an interactive template creation mode wherein an image representative of at least a portion of said document is displayed on the display and the input device conveys instructions from the user to create and store in said system a template which identifies a location within said document information of at least one field of information required by said at least one application program; and

(3) an automatic mode to match at least a portion of said document information with a template created or compatible with said interactive template creation mode stored in said system to extract fields of information required by the application program and to transfer said fields of information to said at least one application program; and

wherein said system is operative in said interactive mode for locating said at least one field of information required by said at least one application program by blocking portions of said at least a portion of said document information displayed on the display using the input device.

53. A multimode information processing system as set forth in claim 30 or claim 31, wherein said computer is operative in said interactive mode for locating said at least one field of information required by said at least one application program by blocking portions of said at least a portion of said document information displayed on the display using the input device.

United States Patent [19]

Lech et al.

[11] Patent Number: 5,768,416

[45] Date of Patent: Jun. 16, 1998

[54] INFORMATION PROCESSING
METHODOLOGY

[75] Inventors: Robert Lech, Jackson; Mitchell A. Medina, Essex Fells; Catherine B. Elias, Plainsboro, all of N.J.

[73] Assignee: Millennium L.P., Grand Cayman, Cayman Islands

[21] Appl. No.: 487,150

[22] Filed: Jun. 7, 1995

Related U.S. Application Data

[62] Division of Ser. No. 348,224, Nov. 28, 1994, Pat. No. 5,625,465, which is a continuation of Ser. No. 143,135, Oct. 29, 1993, Pat. No. 5,369,508, which is a continuation of Ser. No. 672,865, Mar. 20, 1991, Pat. No. 5,258,855.

[51] Int. Cl. 6 G06F 15/40

[52] U.S. Cl. 382/180; 382/175

[58] Field of Search 395/148, 149; 382/287, 175, 306, 177, 180, 282, 317

[56] References Cited

U.S. PATENT DOCUMENTS

4,760,606	7/1988	Lesnick et al.	382/48
4,802,104	1/1989	Ogiso	364/518
4,931,957	6/1990	Takagi et al.	364/521
5,031,121	7/1991	Iwai et al.	364/523
5,228,100	7/1993	Takeda et al.	382/61
5,367,619	11/1994	Dipaolo	395/149
5,404,294	4/1995	Kamik	364/419.1

5,416,849	5/1995	Huang	382/173
5,448,738	9/1995	Good et al.	395/700
5,452,379	9/1995	Poor	382/317
5,506,697	4/1996	Li et al.	358/448

OTHER PUBLICATIONS

Que's Computer User's Dictionary, 2nd Ed., Bryan Pfaffenberger (author); 1991; p. 144.

Primary Examiner—Edward L. Coles, Sr.

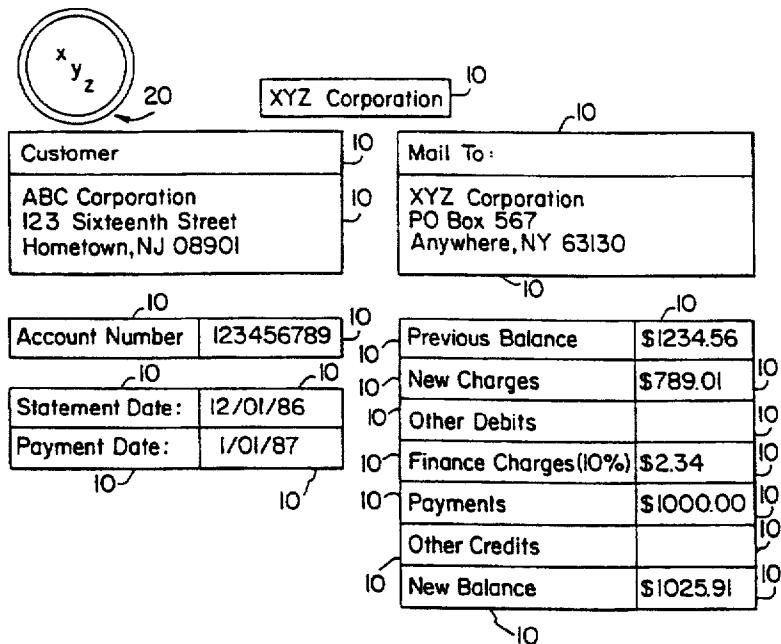
Assistant Examiner—Stephen Brinich

Attorney, Agent, or Firm—Foley & Lardner

[57] ABSTRACT

An information processing methodology gives rise to an application program interface which includes an automated digitizing unit, such as a scanner, which inputs information from a diversity of hard copy documents and stores information from the hard copy documents into a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which designate portions of the stored document information required by a particular application program. The selected stored document information is then placed into the transmission format required by a particular application program in accordance with transmission format instructions. After the information has been transmission formatted, the information is transmitted to the application program. In one operational mode, the interface interactively prompts the user to identify, on a display, portions of the hard copy documents containing information used in application programs or for storage.

66 Claims, 15 Drawing Sheets



Winter Sale In Effect Throughout January!

U.S. Patent

Jun. 16, 1998

Sheet 1 of 15

5,768,416

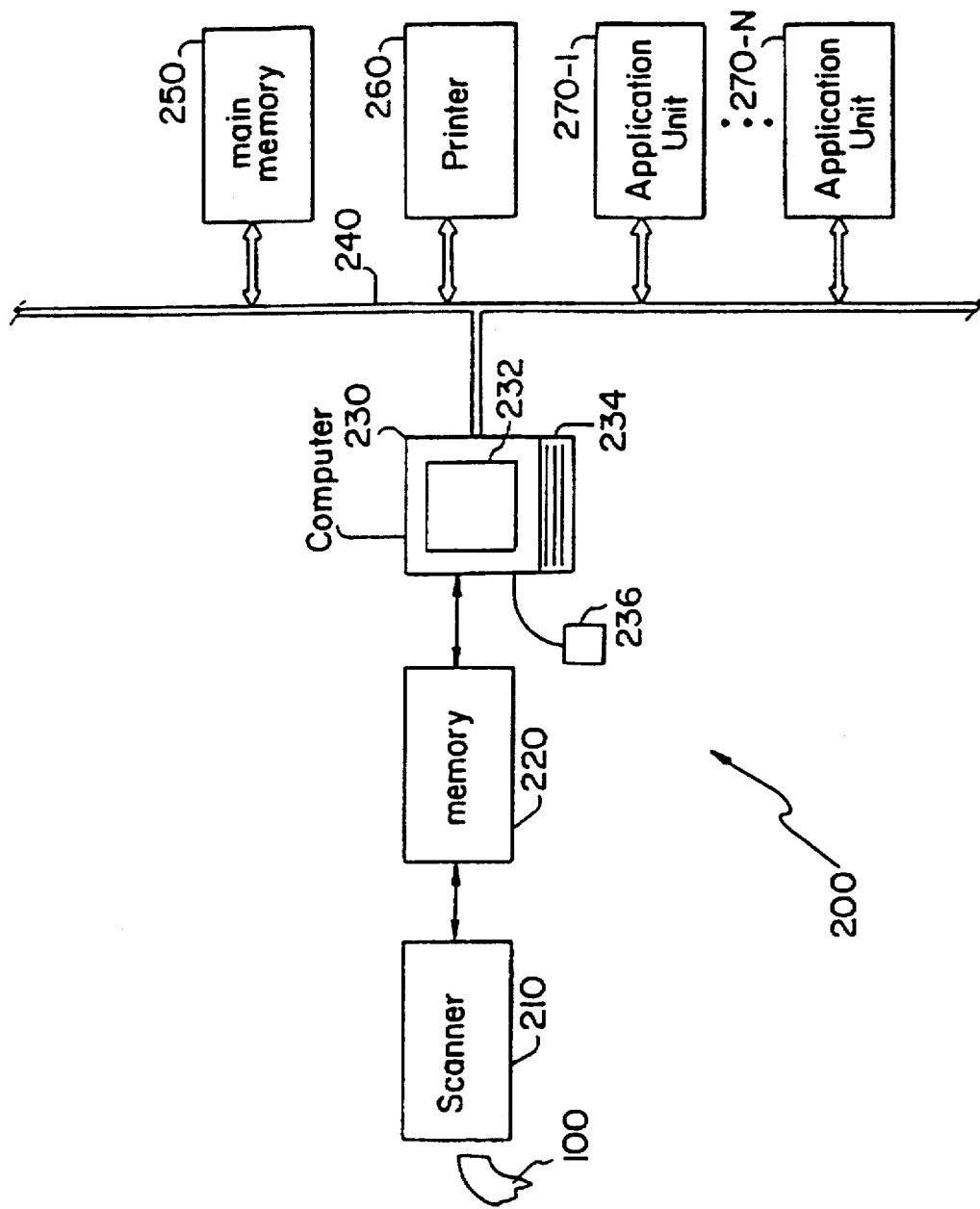


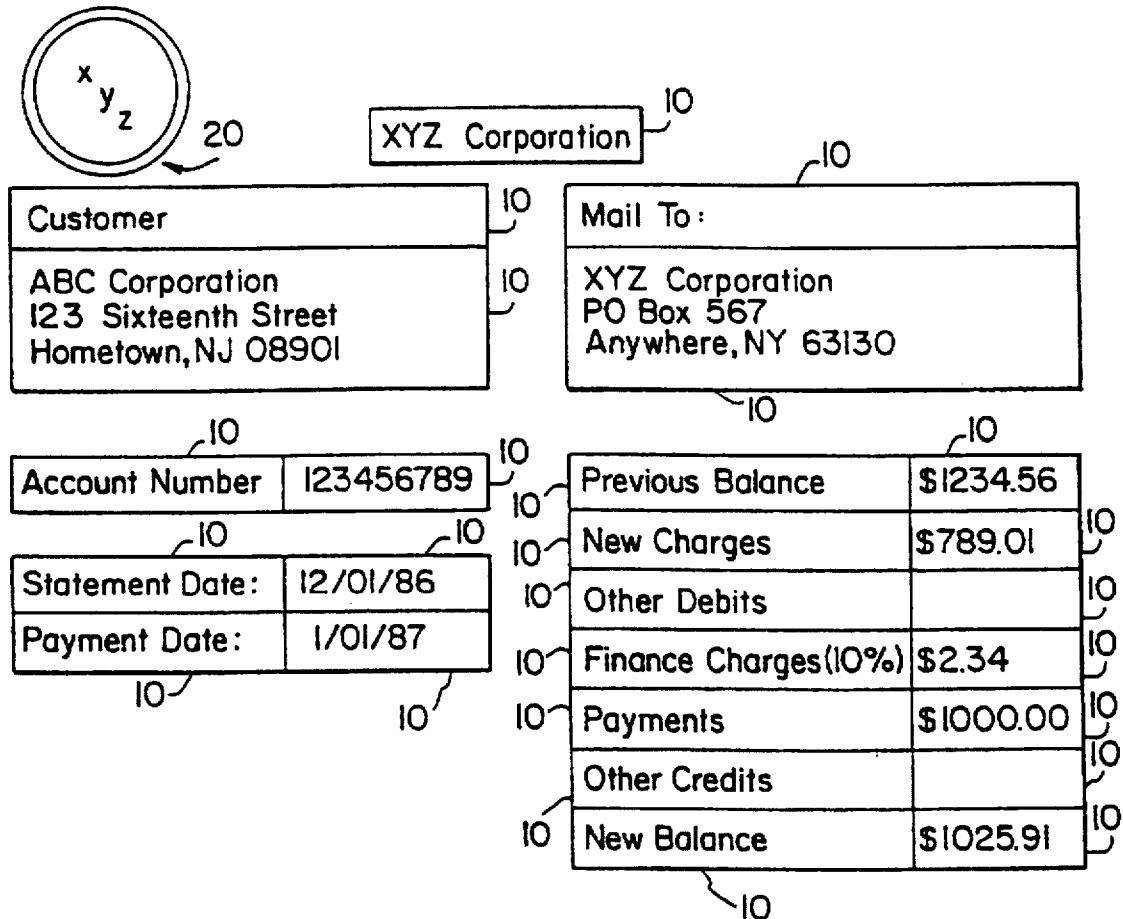
FIG. 1

U.S. Patent

Jun. 16, 1998

Sheet 2 of 15

5,768,416



Winter Sale In Effect Throughout January!

30

FIG. 2

U.S. Patent

Jun. 16, 1998

Sheet 3 of 15

5,768,416

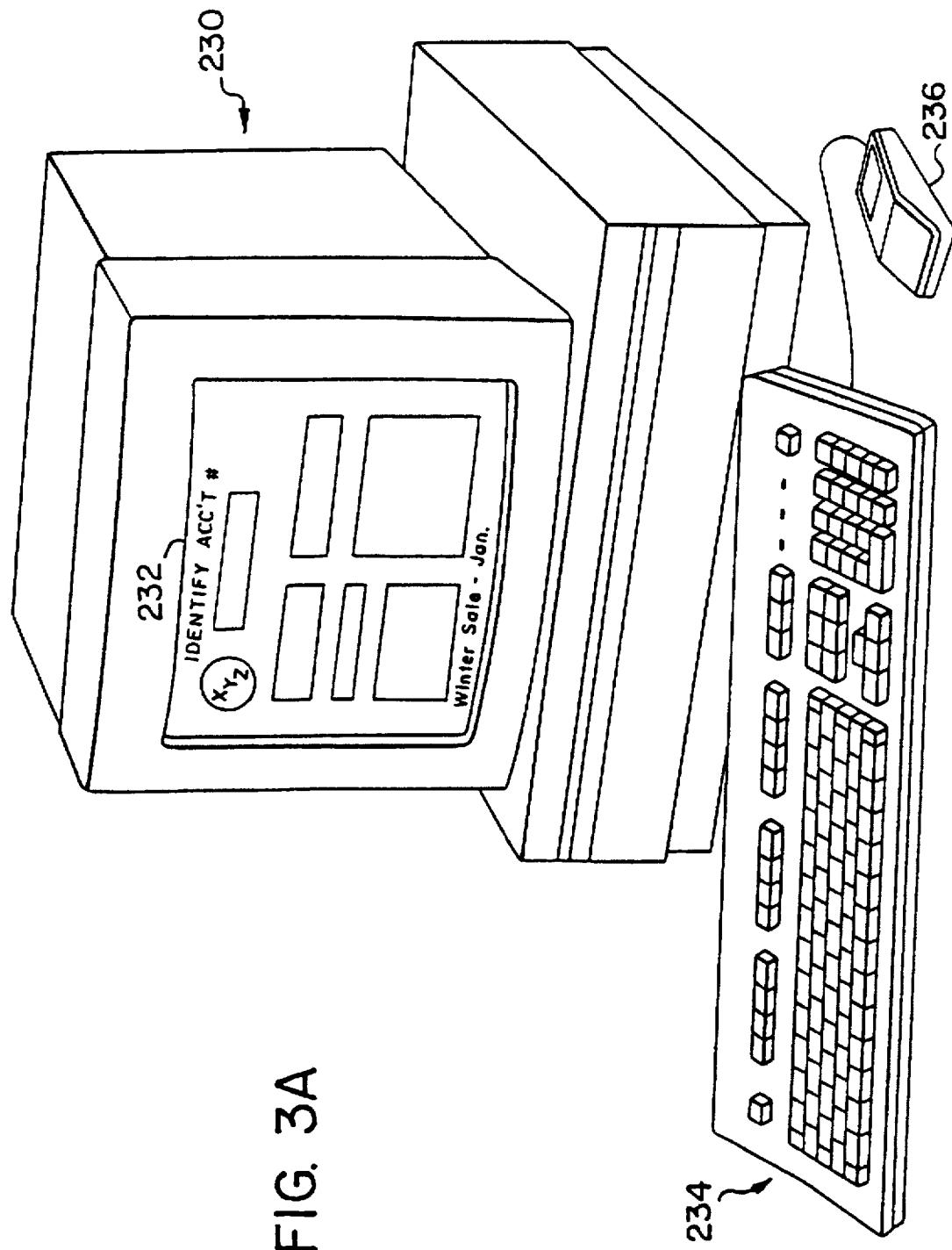


FIG. 3A

U.S. Patent

Jun. 16, 1998

Sheet 4 of 15

5,768,416

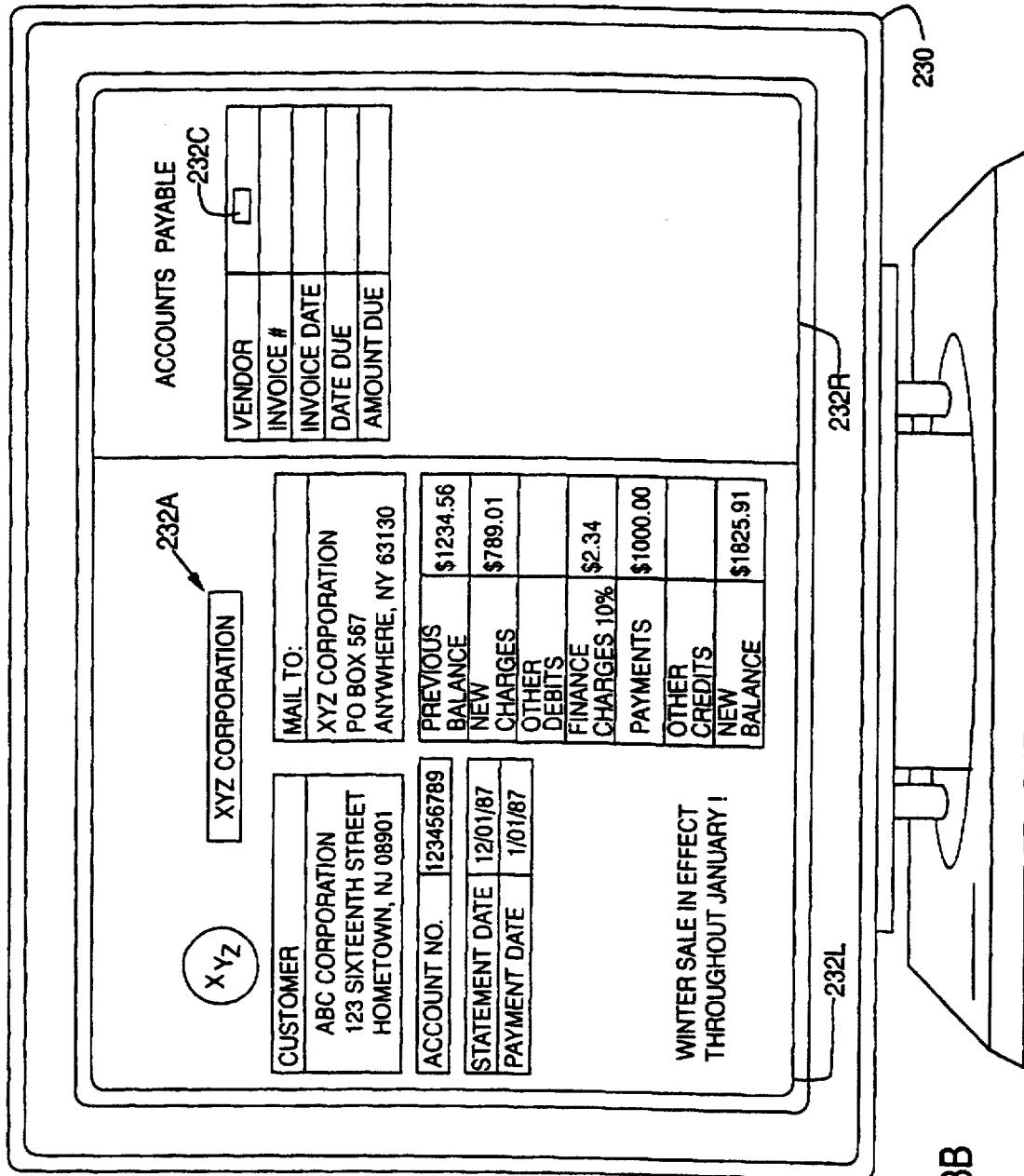


FIG. 3B

U.S. Patent

Jun. 16, 1998

Sheet 5 of 15

5,768,416

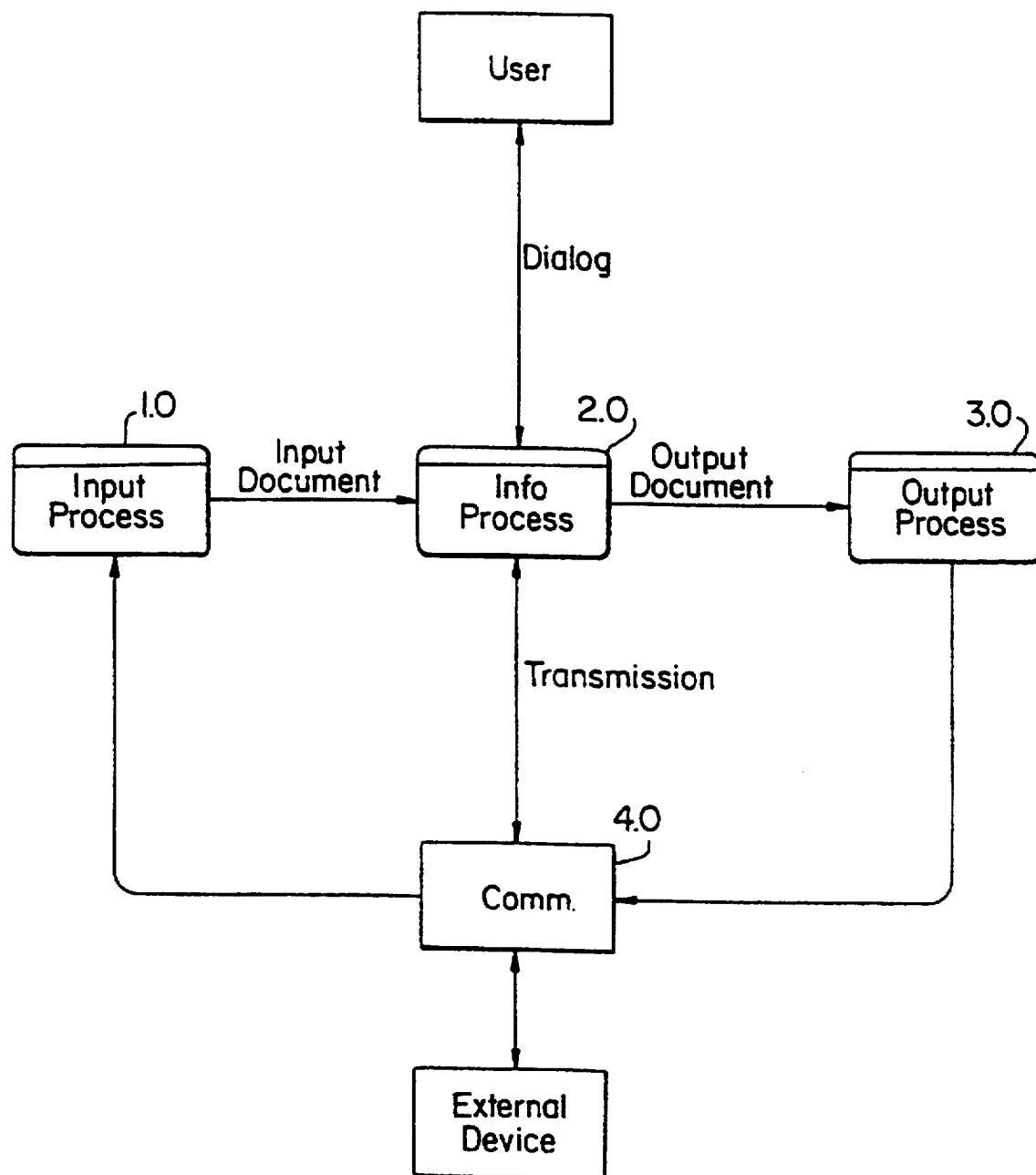


FIG. 4

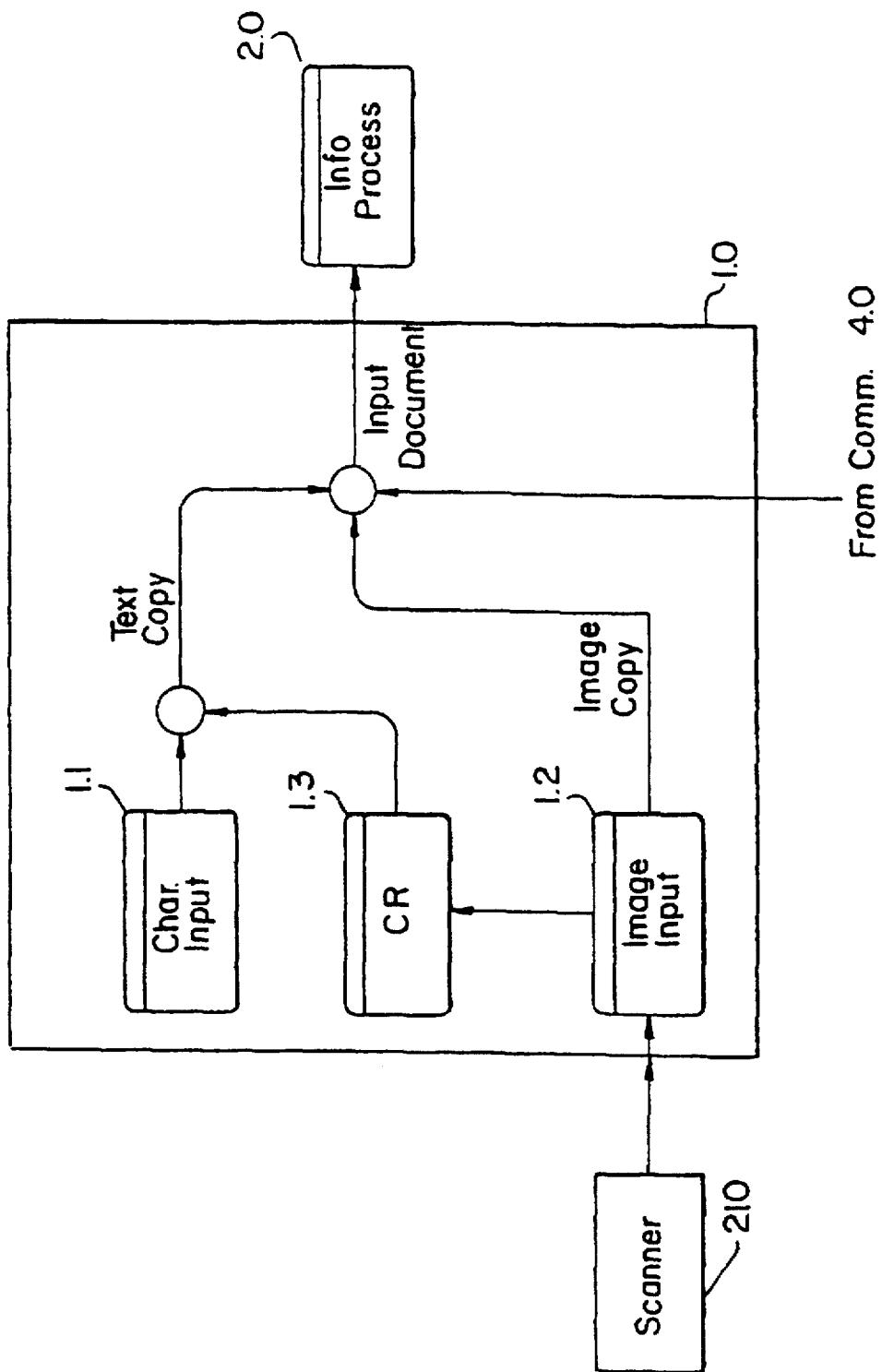


FIG. 5

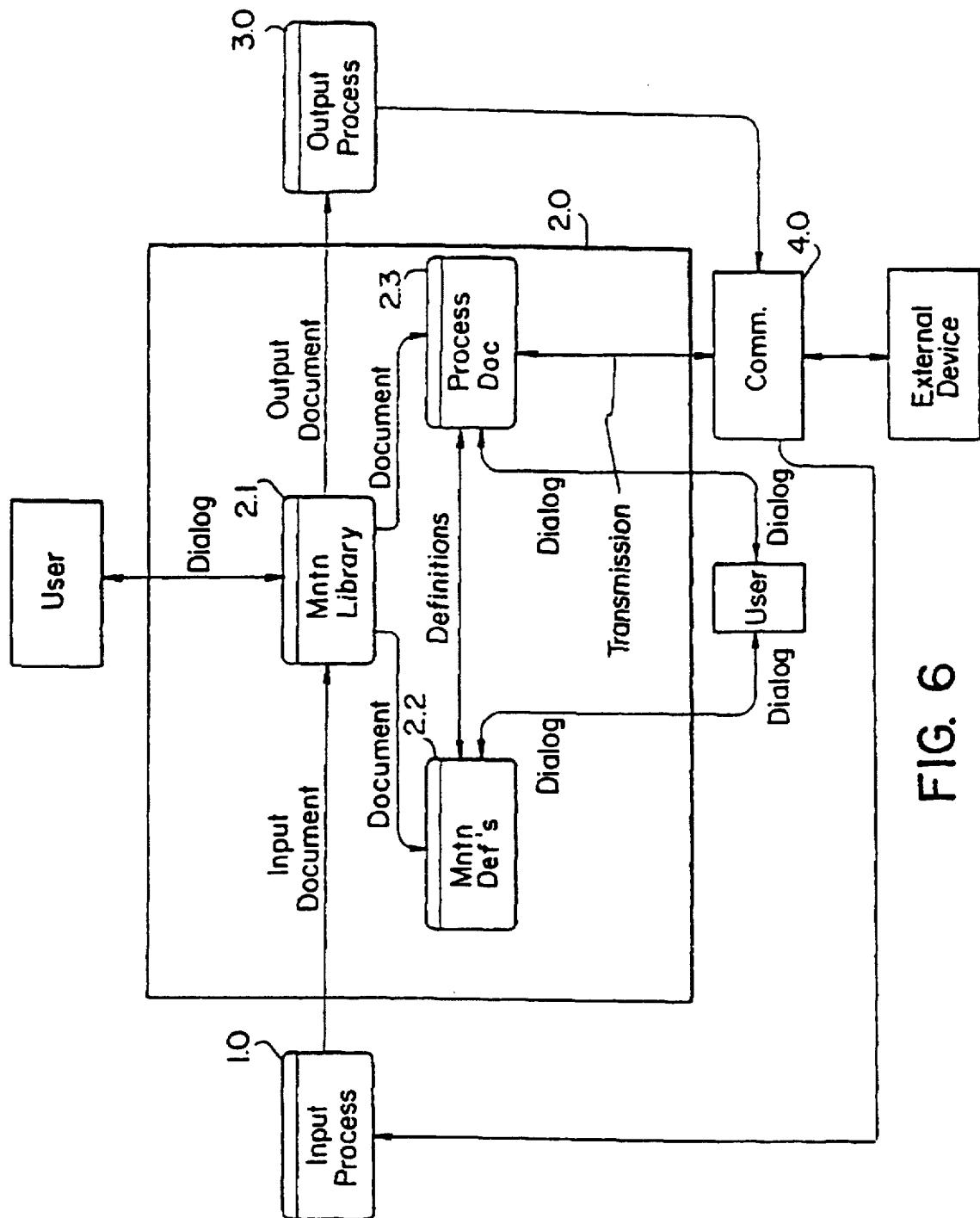


FIG. 6

U.S. Patent

Jun. 16, 1998

Sheet 8 of 15

5,768,416

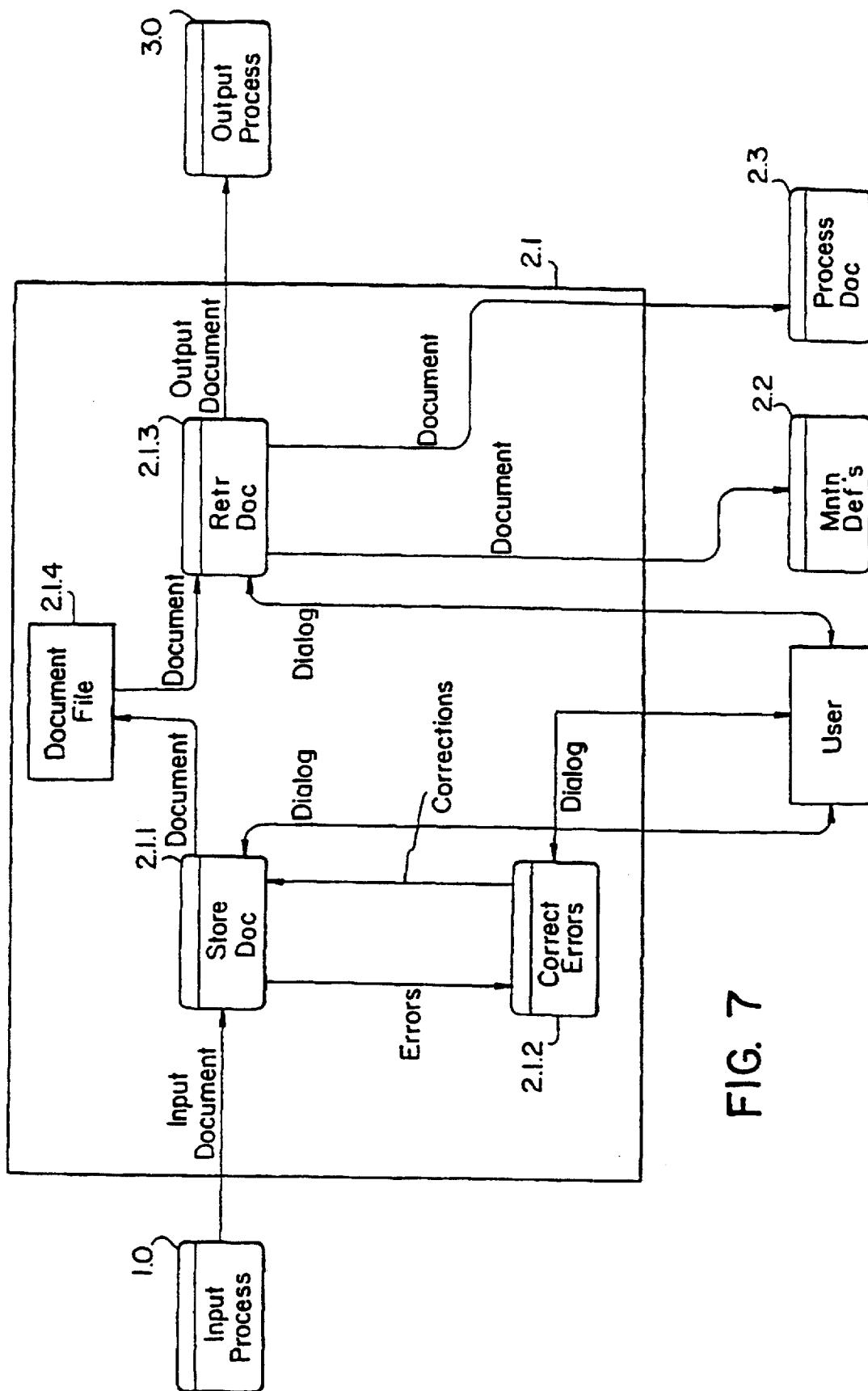


FIG. 7

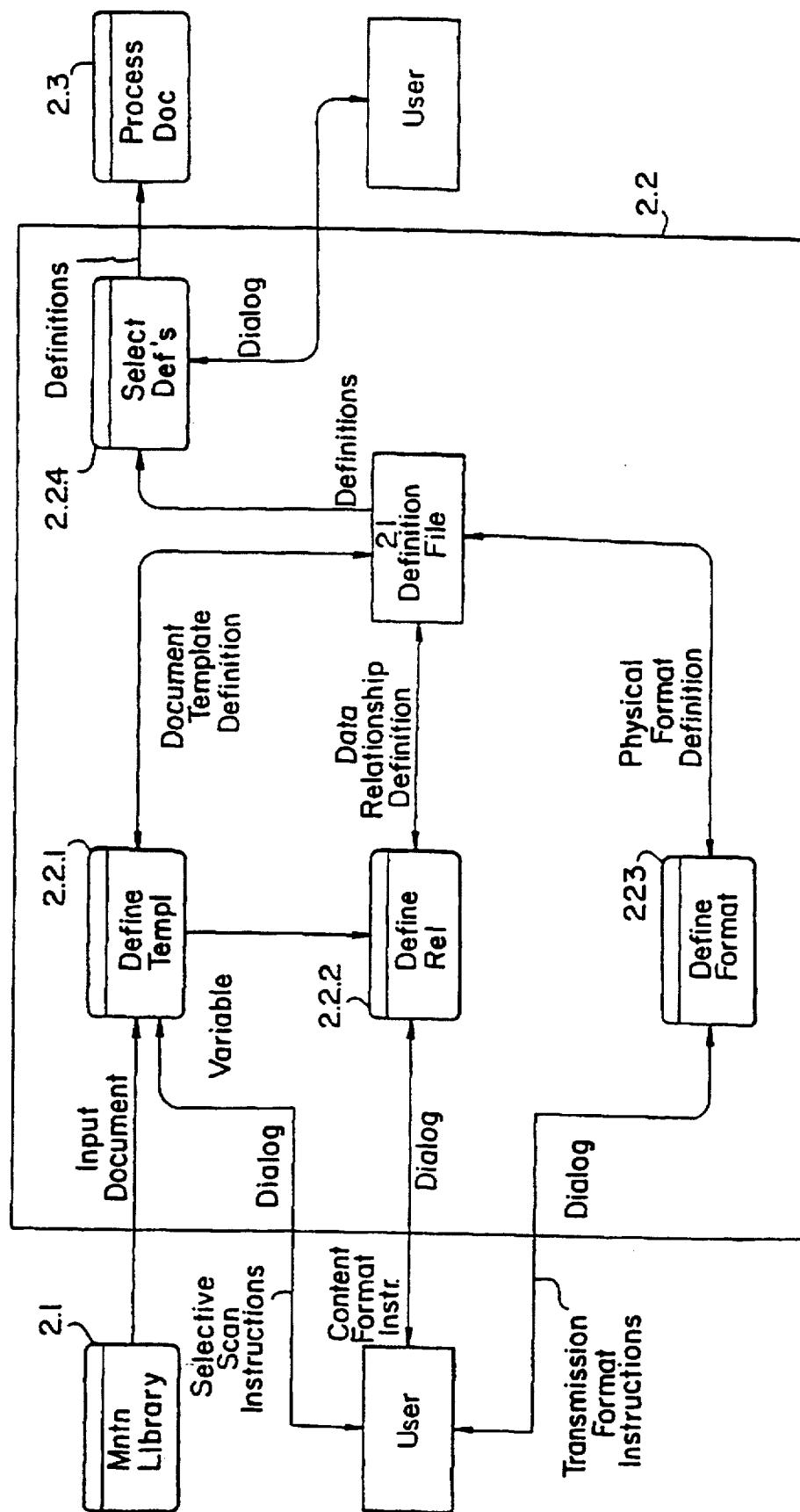


FIG. 8

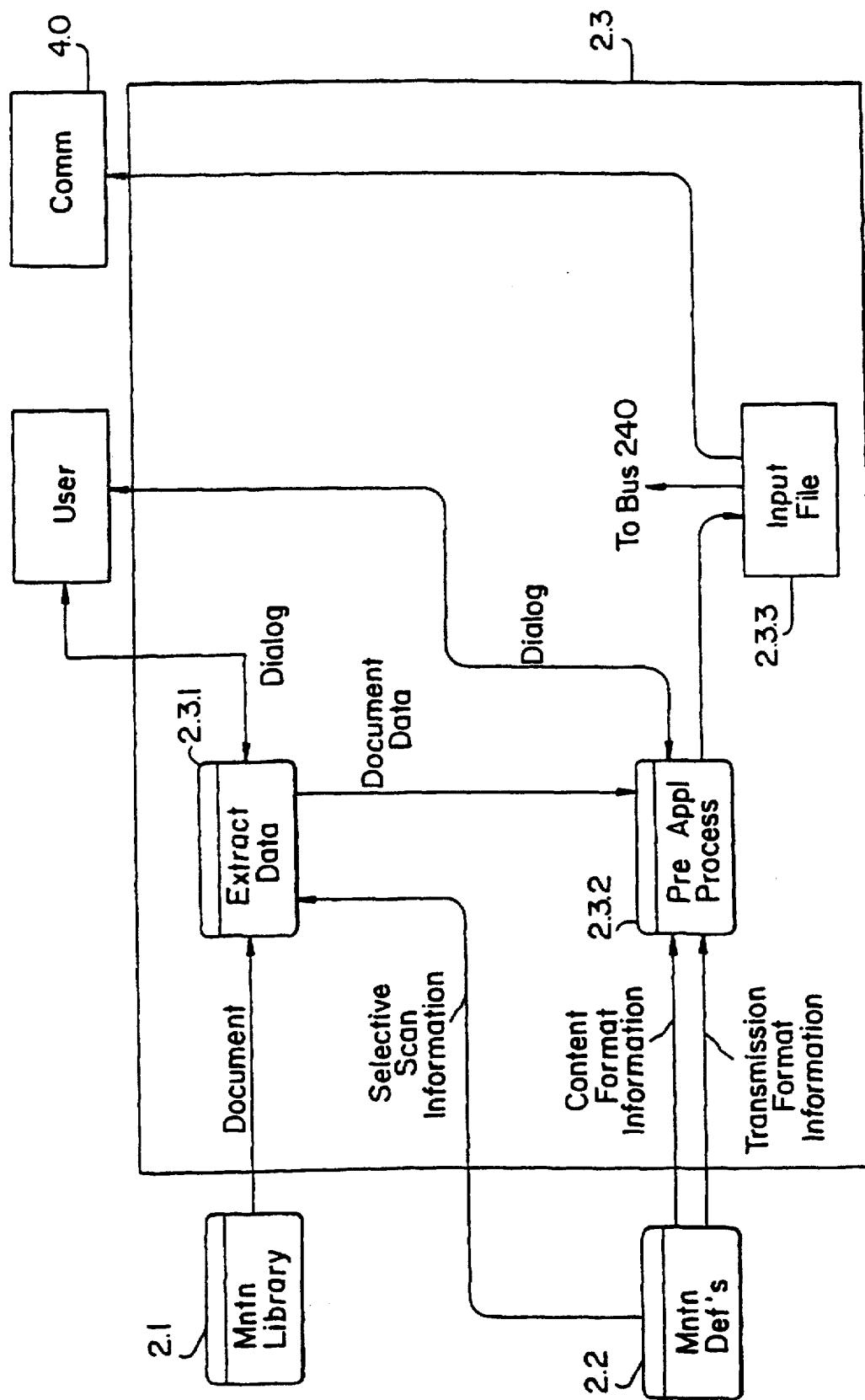


FIG. 9

U.S. Patent

Jun. 16, 1998

Sheet 11 of 15

5,768,416

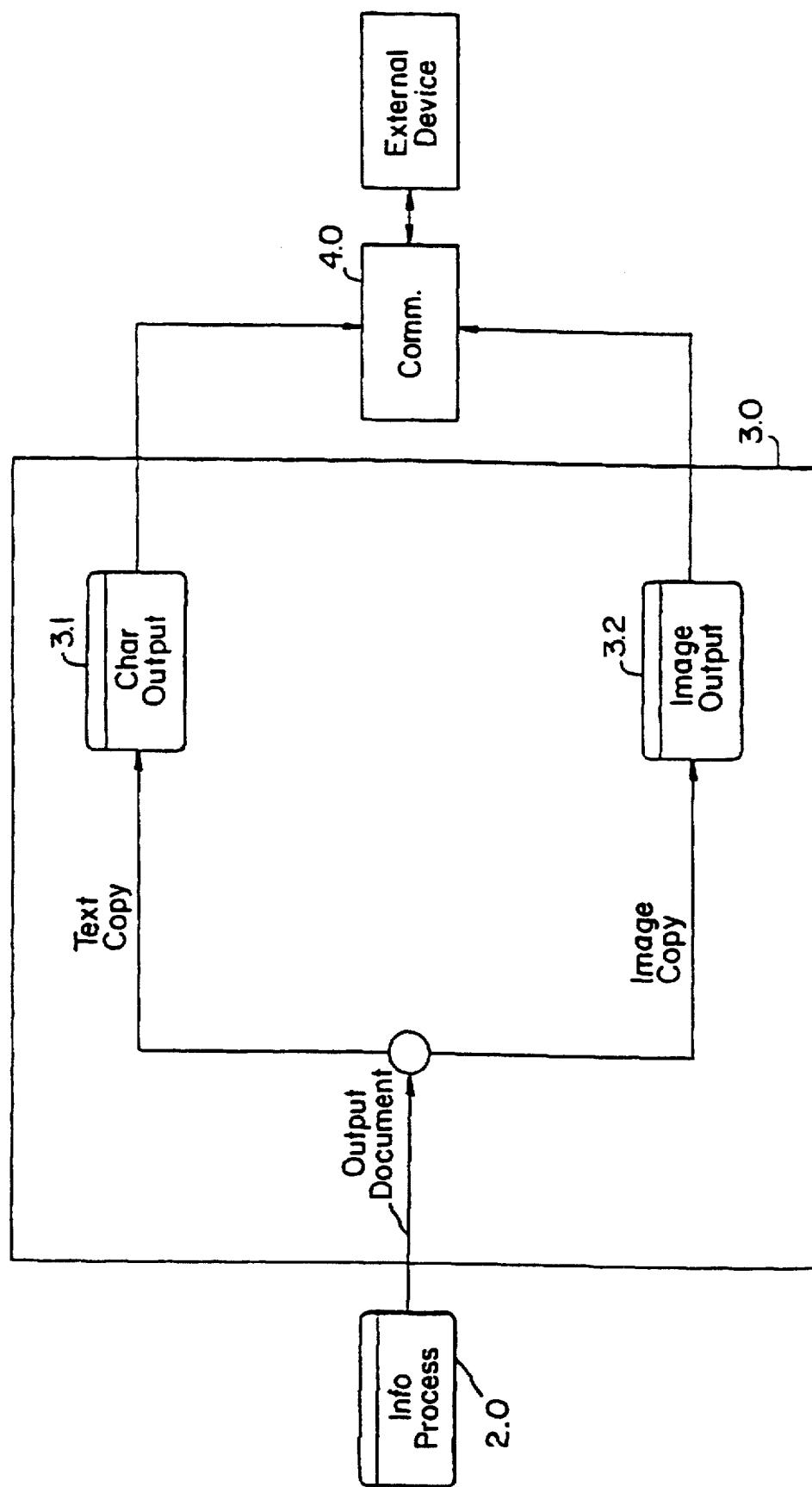


FIG. 10

FIG. 11

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Heading 2	2
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
Other Credits	
New Balance	\$1025.91

U.S. Patent

Jun. 16, 1998

Sheet 13 of 15

5,768,416**FIG. 12A**

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
New Balance	\$1025.91

FIG. 12B

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

FIG. 12C

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

U.S. Patent

Jun. 16, 1998

Sheet 14 of 15

5,768,416**FIG. 13A**

```
>1>1>1>"XYZ Corporation"  
>2>2>25>+123456789>  
>3>2>1>D12/01/86>  
>4>2>11>D12/15/86>  
>5>2>21>D01/01/87>  
>6>10>25>$1234.56>  
>7>11>25>$789.01>  
>8>13>25>$2.34>  
>9>14>25>$1000.00>  
>10>16>25>$1025.91>
```

FIG. 13B

```
>1>1>1>$1234.56>  
>2>2>1>"XYZ Corporation"  
>3>3>1>"PO Box 567"  
>4>4>1>"Anywhere, NY 63130"
```

FIG. 13C

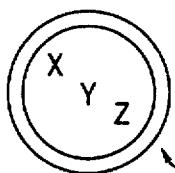
```
/1/1/1/$1234.56//  
/2/2/1/*XYZ Corporation*  
/3/3/1/*PO Box 567*  
/4/4/1/*Anywhere, NY 63130*
```

U.S. Patent

Jun. 16, 1998

Sheet 15 of 15

5,768,416

FIG. 14

XYZ Corporation

Customer

ABC Corporation
123 Sixteenth Street
Hometown, NJ 88981

Mail To:

△ XYZ Corporation
PO Box 567
Anywhere, NY 63130 △

Account Number 123456789

Previous Balance \$1234.56

Statement Date: 12/01/86 New Charges \$789.01

Payment Date: 1/01/87

Other Debits

Finance Charges (10%) \$2.34

Payments \$1000.00

WINTER SALE IN EFFECT THROUGHOUT JANUARY:

Other Credits

New Balance \$1025.91

30

INFORMATION PROCESSING METHODOLOGY

This application is a divisional of application Ser. No. 08/348,224, filed Nov. 28, 1994, now U.S. Pat. No. 5,625,465, in turn a continuation application of Ser. No. 08/143,135, filed Oct. 29, 1993, now U.S. Pat. No. 5,369,508, and which is in turn a continuation of application Ser. No. 07/672,865, filed Mar. 20, 1991, now U.S. Pat. No. 5,258,855.

BACKGROUND OF THE INVENTION

The invention is directed to a system for efficiently processing information originating from hard copy documents. More specifically, the invention is directed to a hard copy document to application program interface which minimizes the need to manually process hard copy documents.

In the past, information contained on hard copy documents was manually entered into a computer via the input controller of a particular computer. The original document was then filed away for future reference. Automatic input of data was limited to the input of Magnetic Ink Character Recognition (MICR) data and to Optical Character Recognition (OCR) data. This fixed-position data was forwarded directly to a dedicated computer application specifically designed to accommodate the input format. In more recent years, typewritten text has been mechanically inputted into a computer via a text file. Examples of this latter type of system are word processors and photo-typesetters.

These conventional systems have limitations which decrease the efficiency of processing information from a hard copy document. For example, the systems discussed above are limited in their application to MICR, OCR, or typewritten data. Parsing and processing data is limited to the particular requirements of the particular computer application which requires the input data. In addition, in these conventional systems, the actual hard copy document must be retained for future reference at great expense.

In a sophisticated computer network, different users may require different portions of the information contained on a hard copy document. For example, if the hard copy document is an invoice returned with payment of a bill, the accounting department may need all of the monetary information contained on the bill while the mailroom may need only customer address information, to update a customer's address. Therefore, there is a need for a system in which specific information from a hard copy document can be selectively distributed to various users.

Another problem with conventional systems is that users, even within the same company, may require that the information extracted from a hard copy document be transmitted to a particular application program in a specific transmission format. For example, one department in a company may use a particular application program which must receive information using a particular character as a delimiter and other departments may require the information in a different format using different delimiters.

Another problem, particularly for small businesses, is that current systems can not efficiently accommodate the inputting of information from a diversity of hard copy documents. A large business which receives many forms in the same format can afford a system which inputs a high volume of information in that format into memory. For example, it is cost-effective for a bank which processes hundreds of thousands of checks a month to buy a dedicated machine which

can read information off of checks having a rigidly defined, or fixed, format. However, as the diversity of forms received by a business increases relative to the number of forms that must be processed, it becomes less cost-effective to design a dedicated machine for processing each type of form format. This problem is particularly significant in small businesses which may, for example, receive fifty invoices a month, all in different, non-fixed, formats. It is frequently not cost-effective for a small business to design dedicated systems for inputting information in each of these various formats. This leaves a small business with no other practical alternative than to manually input the information off of each invoice each month.

SUMMARY OF THE INVENTION

It is an object of the invention, therefore, to provide an application program interface which allows a user to select specific portions of information extracted from a diversity of hard copy documents and allows the user to direct portions of this information to several different users in accordance with the needs of the particular user.

It is also an object of the invention to provide a cost-effective system for inputting hard copy documents which can accommodate hard copy documents in a diversity of formats.

It is another object of the invention to provide an application program interface which allows a user to put information, which is to be transmitted, into a particular transmission format, based upon the needs of the receiver of the information.

It is a further object of the invention to provide an application program interface which will allow the extraction, selection, formatting, routing, and storage of information from a hard copy document in a comprehensive manner such that the hard copy document itself need not be retained.

It is another object of the invention to provide a system which reduces the amount of manual labor required to process information originating from a hard copy document.

A further object of the invention is to reduce the time required to process information originating from a hard copy document so that a higher volume of transactions involving hard copy documents can be processed.

The invention provides an application program interface which inputs a diversity of hard copy documents using an automated digitizing unit and which stores information from the hard copy documents in a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which define portions of the stored document information required by a particular application unit. Selected stored document information is then formatted into the transmission format used by the particular application program based on transmission format instructions. The transmission formatted selected stored document information is then transmitted to the particular application program. The hard copy documents may contain textual information or image information or both.

The interface operates in three different modes.

In a first mode, the interface extracts all of the information from hard copy documents and stores this information in memory. Parsing of various portions of the extracted information is performed in accordance with content instructions.

In a second mode, the user operates interactively with the interface by use of a display and an input device, such as a

mouse. In this second mode, a hard copy document is inputted and displayed on the display. The interface then prompts the user to identify the location of various information. For example, the interface can ask the user to identify the location of address information on the hard copy document. In response, the user positions the mouse to identify address information using a cursor. The identified information is then stored as address information in memory. Subsequently, the interface again prompts the user to identify other pieces of information, which are then stored in the appropriate locations in memory. This process proceeds until all of the information which is desired to be extracted off of the hard copy document is stored in memory.

In a third mode of operation, selected portions of information are extracted off of hard copy documents in accordance with predetermined location information which has been specified by the user. For example, the user can define a template which specifies the location of information on hard copy documents. Templates can be formed in conjunction with second mode operation. Alternatively, the user can instruct the interface to search hard copy documents for a particular character or symbol, located on the hard copy documents. The information desired to be extracted off of the hard copy documents is specified relative to the location of this character or symbol.

The interface can also prompt or receive from an applications program or another information processing system, required information, content instructions, and format instructions.

Other objects, features, and advantages of the invention will be apparent from the following detailed description of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be described in further detail below with reference to the accompanying drawings, in which:

FIG. 1 illustrates hardware for implementing a preferred embodiment of the instant invention;

FIG. 2 illustrates an example of a hard copy document containing information to be processed by the instant invention;

FIGS. 3A and 3B are enlarged views of the computer of FIG. 1 used to explain how the invention interactively prompts a user to identify information;

FIG. 4 is an overall data flow diagram for the FIG. 1 preferred embodiment;

FIG. 5 is a detailed input data flow diagram for the FIG. 1 preferred embodiment;

FIG. 6 is a detailed information processing data flow diagram for the FIG. 1 preferred embodiment;

FIG. 7 is a more detailed information processing data flow diagram for the maintain library module of FIG. 6;

FIG. 8 is a more detailed information processing data flow diagram for the maintain definitions module of FIG. 6;

FIG. 9 is a more detailed information processing data flow diagram for the process document module of FIG. 6;

FIG. 10 is a detailed output data flow diagram for the FIG. 1 preferred embodiment;

FIG. 11 lists data corresponding to the hard copy document of FIG. 2;

FIGS. 12A, 12B, and 12C illustrate examples of data which can be selected from the extracted data of FIG. 11 in accordance with content instructions;

FIGS. 13A, 13B, and 13C illustrate examples of the data of FIGS. 12A, 12B, and 12C formatted in accordance with various transmission format instructions to form input files; and

FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Hardware

The invention provides an interface between information originating from a hard copy document and a computer application unit which uses the information. The computer application unit can be a particular computer application program or a device which is controlled in accordance with instructions or information from the hard copy document.

10 The invention also allows storing a copy of the hard copy document in a memory and retrieving the copy of the hard copy document. By providing a comprehensive and integrated system which can accommodate almost all of the possible uses of information contained on a hard copy document, the instant invention allows for a paperless office.

15 The invention includes hardware and software necessary to extract, retrieve, and process information from the hard copy document. A copy of the actual image of the hard copy document is stored in memory. Textual information extracted from the hard copy document is also stored in memory. Textual information is information, such as alphanumeric characters, which is recognized on the hard copy document and which is stored in a form which corresponds to the particular recognized character. For example, the extracted characters can be stored in the ASCII format in an electronic memory.

20 The user can have all of the information extracted from the hard copy document and stored in memory. Alternatively, the interface can interactively prompt the user to identify specific pieces of information for storage. The interface can also extract specific pieces of information using a predefined template. The interface can also prompt or receive from another information processing system or an applications program desired information, content instructions, and format instructions.

25 The instant invention also provides for parsing information extracted from the hard copy document and for directing this parsed information to specific users or application programs as an input file.

30 The invention also permits the user to define the transmission format of the input file for a particular computer application unit.

35 FIG. 1 illustrates hardware for implementing a preferred embodiment of a hard copy document to application program interface according to the instant invention. The interface 200 processes information extracted off of hard copy document 100 and provides information to application units 270 in a form required by each particular application unit. The interface extracts information off of a hard copy document 100 utilizing a scanner 210. The scanner 210 can be any type of scanner which extracts information off of hard copy documents, for example, an optical Reader.

40 The scanned information is stored in a scanner memory 220 or in main memory 250, as will be described in greater detail below. If main memory 250 or another memory is available to store the scanned information, then scanner memory 220 can be omitted.

45 The information from scanner memory 220 or main memory 250 is transmitted to computer 230. In the preferred embodiment, computer 230 includes a display 232, a keyboard 234, and a mouse 236. The display 232 displays an

image of the hard copy document itself and/or information necessary to process the information extracted off of the hard copy document.

The computer 230 is used to select portions of the stored document information contained in memory in accordance with content instructions which define portions of the stored document information required by an application unit. These content instructions may be provided by the application program. Alternatively, the content instructions can be inputted via an input device such as a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

The computer 230 is also used to format selected stored document information into the transmission format used by an application unit based on transmission format instructions. The transmission format instructions may be provided by the application program. Alternatively, the transmission format instructions can be inputted via a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

Thus, the computer 230 is used to generate an input file for a particular application unit. The computer 230 is connected to scanner memory 220, main, or permanent, memory 250, a printer 260, and application units 270, via a bus 240. Although FIG. 1 illustrates use of a bus to connect components together, it is understood that any routing or connecting link, implemented in hardware or software or both, can be employed instead of, or in addition to, a bus. Instructions to or in the computer 230 control the main memory 250, the printer 260, the application units 270, and the bus 240. Instructions to or in computer 230 can also control exchanges of information with scanner memory 220.

When the computer 230 generates an input file for a particular document, the computer 230 can send this input file directly to an application unit or can store this input file in the main memory 250 until required by an application unit. The main memory 250 may also optionally store a copy of the image information for the hard copy document and the textual information for the hard copy document. Thus, the image information and textual information from the hard copy document can be retrieved and printed out on printer 260. In addition, image and textual information stored in scanner memory 220 or in main memory 250 can be used to form additional input files at the time of input or at a later time, based on content instructions and transmission format instructions. Thus, the invention can, at the discretion of the user, eliminate the need to retain copies of hard copy documents, permitting a paperless office.

The application units 270 include particular application programs and devices which are controlled in accordance with information contained on hard copy document 100.

FIG. 2 illustrates an example of a hard copy document 100 which contains information to be processed by the instant invention. The document illustrated in FIG. 2 is a bill from XYZ Corporation to customer ABC Corporation. FIG. 2 is only an example of a type of document that can be processed by the instant invention.

In a first operational mode, the scanner 210 stores all of the information extracted off of hard copy document 100 in the scanner memory 220 or, alternatively, in main memory 250. The extracted information is stored in two forms. The actual image of the hard copy document 100 is stored as image information in the scanner memory 220. In addition, the scanner memory 220 stores textual information recognized on the hard copy document 100 by, for example, employing standard character recognition software. In the preferred embodiment, the textual information is stored in

ASCII format. The scanner memory 220 can be, for example, an electronic, magnetic, or optical memory.

FIG. 3A illustrates an enlarged view of the computer 230 of FIG. 1. This view will be used to describe a second mode of operation. In this second mode of operation, the hard copy document 100 is scanned and a copy of the document 100 is displayed on display 232 of computer 230, based on the contents of information temporarily stored in scanner memory 220. After the document is displayed on display 232, the computer 230 interactively prompts the user to identify the location of specific pieces of information on the hard copy document. In the FIG. 3A illustration, this prompt message is indicated as the message beginning with the arrow.

For example, the prompt message can ask the user to identify the location of account number information on the hard copy document. The user then uses an input device, such as keyboard 234 or mouse 236 or a touch screen, notepad, voice recognition device, or other input device to position a cursor on the display to identify the location of the information requested by the prompt message. For example, the cursor could be used to define a block (which could be highlighted) containing the requested information, followed by a mouse "enter" click. In this example, the user would move the mouse to identify the location of the account number information contained on the hard copy document 100. The computer 230 then stores the information which has been identified by the user as account number information in the appropriate address or subfile or as the appropriate variable or parameter, or data field, in memory. The computer then prompts the user to identify the location of other information on the hard copy document, such as, statement date information. The process proceeds until all of the desired information has been stored into the appropriate locations in memory.

FIG. 3B illustrates a variation of the second mode for interactively prompting the user for information. In FIG. 3B, the display is split into two portions. A left-hand portion 232L displays the image of the hard copy document and a right-hand portion 232R displays the required application program information. For example, in FIG. 3B, portion 232R displays a spreadsheet used by an application program. While observing the split display, the user can input instructions to associate specific pieces of information on the hard copy document (for example, the vendor name indicated by the mouse arrow 232A) with particular subfiles in memory (for example, the vendor field next to which the cursor 232C appears), using a mouse or other input device(s) or both. The split display also allows the user to generate content format instructions while observing the information required for a particular application program on the right-hand portion.

These second modes of operation are efficient for small businesses which receive a small number of a wide variety of invoices, since the user does not necessarily have to store all of the information that appears on the hard copy document. A further advantage is that data input is quicker, easier, and more accurate than with previous keyboard methodology. In addition, by specifying the location on the hard copy document of information, the user may optionally create a template, to be described in further detail below, for each different type of invoice. This template is stored for future use when another hard copy document in the same format is received.

More specifically, instructions from computer 230 can direct the scanner 210 and scanner memory 220, and/or main

memory 250, to scan and/or store only specific portions of hard copy document 100. After the interactive prompts required to obtain information for a desired application program, the unused information stored in scanner memory 220 or 250 can be erased. Further, scanning of a second identical document can be limited to only those portions of the document which contain needed information.

More specifically, in FIG. 2, the lines 10 drawn around certain portions of the document represent the areas which the user has previously identified as the portions of a document to be extracted by the scanner 210 and stored in scanner memory 220 and/or main memory 250. Since the logo 20 and the message 30 have not been identified as an area to be scanned and stored, these areas are not scanned and stored in subsequent documents. Since the user has previously associated each of the areas 10 with a specific subfile of information, e.g., the account number, the scanned information is stored in memory locations corresponding to that subfile.

Data Processing

FIGS. 4-10 illustrate the flow of data in the FIG. 1 preferred embodiment. FIG. 4 illustrates the overall data flow for the FIG. 1 preferred embodiment. The preferred embodiment includes an input process module 1.0, an information processing module 2.0, and an output processing module 3.0. The information processing module 2.0 is equipped to receive instructions from and transmit information to a user. The information processing module 2.0 can also transmit to and receive information from a remote external device through communication interface 4.0. Input process module 1.0 and output processing module 3.0 can also access communication interface 4.0. A module is implemented in hardware, software, or a combination of hardware and software. The specific implementation for a particular business application depends upon a variety of factors, for example, the relative costs of hardware and software implemented systems, the frequency with which a user will want to expand or modify the system, and the like.

FIG. 5 is a more detailed diagram of the input process module 1.0 of FIG. 4. The input process module 1.0 includes a character input module 1.1, an image input module 1.2, and, in the preferred embodiment, a character recognition device 1.3. The character input module inputs textual information, such as alphanumeric characters, from an input device such as keyboard 234. The image input module 1.2 inputs image information, for example, a digitized image of the actual appearance of hard copy document 100. Textual information can include textual input from an input device such as keyboard 234 and textual information extracted from the document by character recognition device 1.3. Both types of information comprise an input document which is transmitted to information processing module 2.0. In the FIG. 1 preferred embodiment, the processing performed by input process module 1.0 occurs in scanner memory 220, computer 230, and main memory 250.

FIG. 6 illustrates information processing data flow for the FIG. 1 preferred embodiment, that is, FIG. 6 illustrates data flow in the information processing module 2.0.

The information processing module 2.0 includes a maintain library module 2.1, to be described in further detail below in conjunction with FIG. 7, a maintain definitions module 2.2, to be described in further detail below in conjunction with FIG. 8, and a process document module 2.3 to be described in further detail below in conjunction with FIG. 9.

The information processing module 2.0 is the module which coordinates and drives the entire system. In the preferred embodiment, the information processing module 2.0 is implemented primarily by computer 230.

FIG. 7 illustrates information processing data flow in the maintain library module 2.1. The maintain library module 2.1 maintains a library of imaae information, for example, a digitized image representing the actual appearance of the hard copy document, and textual information of the hard copy documents for reference during processing. This library can be incorporated within scanner memory 220, main memory 250, or another independent memory, for example, a RAM disk. The maintain library module 2.1 includes a store document module 2.1.1, a correct errors module 2.1.2, a retrieve document module 2.1.3, and a document file 2.1.4. These modules operate collectively to store, retrieve, and correct document information.

The store document module 2.1.1, prior to routing the document to the document file 2.1.4, may provide information on recognition errors which may have occurred while inputting the document. For example, the store document module 2.1.1 identifies that a character contained on hard copy document 100 was not recognized. The store document module 2.1.1 also optionally causes a copy of the document and its parsing to be displayed on the display 232 for confirmation by the user. The user may utilize this opportunity to identify any errors in the displayed document and, in conjunction with the correct errors module 2.1.2, to revise the document's parsing, if necessary, prior to storage of the document in memory. The module 2.1.1 also provides a facility for the user to name a particular hard copy document for cataloging, storage, and retrieval purposes. After the document is named, the store document module 2.1.1 stores copies of the document in the document file 2.1.4.

The correct errors module 2.1.2 processes instructions from the user to correct errors identified by the store document module 2.1.1 and errors that have been spotted by the user during the confirmation process.

The retrieve document module 2.1.3 permits the user to retrieve a copy of a document previously stored in the document file 2.1.4. As described above, long-term storage is provided by main memory 250, if necessary.

FIG. 8 illustrates a more detailed information processing data flow diagram for the maintain definitions module 2.2 of FIG. 6. The maintain definitions module 2.2 allows the user to define system and document parameters and maintains the definitions of these system and document parameters. The maintain definitions module 2.2 includes a define template module 2.2.1 which allows the user to specify the location of information on the document. This information provided by the user defines a template which is used to extract information off the document and to associate the extracted information with a particular variable or subfile. These templates are illustrated by boxes 10 in the FIG. 2 example of a hard copy document. The maintain definitions module 2.2 can also access templates previously defined by the user and stored in main memory 250. Templates can also be provided as part of software packages developed by program developers.

The maintain definitions module 2.2 also includes a define relationships module 2.2.2. The define relationships module 2.2.2 allows the user to define data relationships, or logical relationships, between pieces of information extracted from the hard copy document. These pieces of information are then used to generate an input file for a selected computer application unit. The user defines these relationships by

content instructions. Alternatively, content instructions to define relationships can be provided by application software. If the user provides these content instructions, the content instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. Examples of content instructions, data, and logical relationships will be described in further detail in conjunction with FIGS. 11 and 12A, 12B, and 12C.

The maintain definitions module 2.2 also includes a define format module 2.2.3. The define format module 2.2.3 allows the user to define transmission formats for an input file which is then transmitted to a selected computer application unit. Selection of the transmission format of the input file is accomplished by the user through use of transmission format instructions. Alternatively, the applications software itself can generate its own transmission format instructions. When the user must specify transmission format instructions, the transmission format instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. A further description of various transmission formats will be provided below in conjunction with FIGS. 12A, 12B, 12C, 13A, 13B, and 13C.

A select definitions module 2.2.4 is also included in the maintain definitions module 2.2. The select definitions module 2.2.4 allows the user to store and select a set of definitions to be used for processing the document. The definitions identify pieces of information on the document by, for example, absolute location, variable location, or relative location, or by proximity to key words and/or symbols. These definitions are described in further detail below by way of an illustrative example.

FIG. 9 illustrates a more detailed information processing data flow diagram for the process document module 2.3. The process document is module 2.3 processes the document after the document has been stored in the system. The process document module 2.3 gathers the appropriate information which has been stored, and creates input file(s) 2.3.3 for the selected application unit. The process document module 2.3 then transmits the input file(s) via bus 240 and/or communication interface 4.0 to an application unit 270, an output device such as printer 260, or to main memory 250.

The process document module 2.3 includes an extract data module 2.3.1. This module extracts data off of the document in accordance with the user's instructions, for example, the user-defined template, or through the interactive mode.

The process document module 2.3 also includes a preapplication process module 2.3.2 which gathers and associates information extracted from the document in accordance with content instructions. This module prompts the user for any additional information required to satisfy the relationships defined by the content instructions. The preapplication process module 2.3.2 also places the selected information into the transmission format defined by the transmission format instructions.

The preapplication process module 2.3.2 also generates the input file 2.3.3 for the selected application in accordance with the appropriate instructions. The input file 2.3.3 is then transmitted to bus 240 and/or communication interface 4.0 for transmission to a particular application unit 270.

FIG. 10 illustrates a detailed output data flow diagram for output module 3.0. Output module 3.0 outputs a textual and/or image copy of the document. In the FIG. 1 preferred embodiment, output module 3.0 is implemented by printer 260, associated software, and associated interface circuitry.

Operation

Examples of operation of a preferred embodiment will now be described.

The user enters the system by providing instructions to the information processing module 2.0. The user then instructs the information processing module 2.0 to conduct maintain library processing, maintain definitions processing, or process document processing.

If the user selects maintain library processing, the user then provides instructions to maintain or modify the document library through the maintain library module 2.1. For example, the user can direct the inputting and storage of a hard copy document 100 or can retrieve and output a document. The user requests inputting of a document through the store document module 2.1.1. The system then prompts the user to specify a storage location for the inputted document. The document is then read in by the input process module 1.0. A textual copy and/or an image copy are stored into the document file 2.1.4. Errors which have occurred during inputting are identified and corrected by the correct errors module 2.1.2 and the user. The corrections are reflected in the document information stored in document file 2.1.4.

The retrieve document module 2.1.3 is used to retrieve and output a document. The system prompts the user to specify the storage location of a document and the type of document copy, for example, a textual or an image copy, to be outputted. The document is then outputted by the output process module 3.0.

If the user initially selected maintain definitions processing, the user would instruct the system to maintain and/or modify parameter definitions through the maintain definitions module 2.2. For example, the user can define and maintain a document template for extracting selected portions of information off of the hard copy document. The user can use the template to extract selected portions of information off of the hard copy document when the document is originally inputted, or alternatively, the user can use the template to identify selected portions of information for extraction off of an image copy of the document. In creating the template, the user identifies pieces of information on the document to be extracted and assigns a variable name, or subfile, to each piece of data.

The location of data to be extracted can be defined in a number of ways other than by use of a template. For example, the user can designate the absolute location of information on the document with respect to a grid overlaid on the document, e.g., always on line 3, starting in column 1. The user can also identify information by specifying the relative location of information to be extracted, e.g., always two lines below the piece of data named "salutation", starting in column 3. The user can also specify the location of information to be extracted by variable location specification. For example, if the hard copy document is a letter, the module would conduct a key word search for the term "Dear Sir.". Wherever this term "Dear Sir." is located, this piece of data would be associated with the variable specified by the user, for example, the variable "salutation." In addition, a defined set of conventional symbols can be used to signify certain recurring data items for the convenience of users of the instant invention. For example, a "@" symbol can be used to delineate the vendor name as follows: "@XYZ Corporation@". Other examples of the use of symbols to delineate information will be described with reference to FIG. 14.

The maintain definitions module 2.2 is also used to maintain data relationships in accordance with content instructions and to maintain input file formats in accordance with transmission format instructions. Relationships are

defined and maintained between pieces of data, specified by, for example, the names of variables, through the define relationships module 2.2.2. The names of pieces of data on the document are retrieved by, for example, the define template module 2.2.1, and are passed to the define relationships module 2.2.2. The user may then provide any additional pieces of data needed to generate an input file for a particular application program or unit, such as an input file line number. The user, the applications software, and/or instructions previously stored in memory then establishes the contents of the input file by defining relationships between pieces of data using content instructions. Specific examples of content instructions will be discussed below in conjunction with FIGS. 11, 12A, 12B, 12C, 13A, 13B, and 13C.

The user and/or the applications software defines and maintains the transmission format of the input file to be used by a particular application program or unit through the define format module 2.2.3 in accordance with transmission format instructions. This is accomplished by defining the parameters to be used by the preapplication process module 2.3.2 in generating an input file. Parameters which would typically be required to generate an input file would include the character type, e.g., text or pixel; delimiters used between pieces of data, e.g., a slash or a semicolon; end of line characters, e.g., a carriage return or a line feed; and end of file characters. Examples of transmission formats will be described in further detail below in conjunction with FIGS. 11, 12A, 12B, 12C, 13A, 13B, and 13C.

If the user initially selected process document processing, the interface will then proceed to process the document through use of the process document module 2.3. For example, the user can extract specific portions of data from an image copy of a document, can generate an input file for transmission to an application program, or can directly process information interactively with an application program.

If the user desires to extract specific portions of data from an image copy of a hard copy document which has already been stored in memory, the user uses the extract data module 2.3.1 to identify a document to be processed. The document is then retrieved by the retrieve document module 2.1.3 and passed to the extract data module 2.3.1. The user can also select parameter definitions through the select definitions module 2.2.4.

The selected document template or parameter definition is passed to the extract data module 2.3.1. The extract data module 2.3.1 extracts pieces of data from the image copy of the document, as defined by the document template definition or the parameter definitions or both. This document data is then passed to preapplication process module 2.3.2.

The interface generates input file(s) 2.3.3 by use of the preapplication process module 2.3.2. The selected data relationship definition, as defined by the content instructions, and the selected record format definitions, as defined by the transmission format instructions, are passed to the preapplication process module 2.3.2. The preapplication process module 2.3.2 assembles the input file in accordance with the content instructions. The preapplication process module 2.3.2 also prompts the user for any additional pieces of data which need to be provided by the user. The input file is converted to the desired transmission format in accordance with the transmission format instructions. This physically formatted data is then stored in the input file 2.3.3.

The user can also use an application program to process information by loading the particular application program

into the computer 230 rather than by sending the input file to a remote application unit 270.

An illustrative example of the processing described above will now be described.

5 The user inputs instructions via keyboard 234 or another input device which indicate that the user desires to input and store a document. The computer 230 then prompts the user for the name of the document. In this example, the user desires to input the document of FIG. 2 and therefore names the document "XYZ Corp. Bill 12/01/86." The computer then prompts the user to feed the hard copy document 100 into the scanner 210. The image of the hard copy document is displayed on display 232. The computer then prompts the user to identify the account number on the document. By use 10 of the mouse 236 or other input device to position a cursor on the display, the user indicates the location of the account number. The account number is then read-in to a subfile named "Account Number." This process proceeds until all of the desired information has been read-in and stored.

15 In this particular example, no errors were encountered while inputting the document. The user then directs that the document be stored for future reference in a document file.

20 Some time later, the user desires to retrieve and output the document and to generate input files based on information from the document. The computer 230 prompts the user for the name of the document and the type of output. The user responds with "XYZ Corp. Bill 12/01/86" for a printed textual copy. The document is then retrieved from the document file and passed to the printer 260 for printing.

25 In order to generate an input file for a specific application program, the user selects the option to define a document template for use when each month's XYZ Corporation bill arrives. Accordingly, the user instructs the system to display a copy of an XYZ Corporation bill on the display 232. The user then identifies pieces of data by absolute locations. That is, the user assigns specific names to information located at specific portions of the document. In this example, the user would input the following information:

30 Vendor-text, line 1, one line, column 1, 80 characters;
Account number-numeric, line 6, one line, column 25, 9 characters;
Statement date-date, line 9, one line, column 25, 8 characters;
35 Payment date-date, line 11, one line, column 25, 8 characters;
Previous balance-currency, line 7, one line, column 75, 9 characters;
New charges-currency, line 8, one line, column 75, 9 characters;
Other debits-currency, line 10, one line, column 75, 9 characters;
Finance charges-currency, line 12, one line, column 75, 9 characters;
50 Payments-currency, line 13, one line, column 75, 9 characters;
Other credits-currency, line 14, one line, column 75, 9 characters;
New balance-currency, line 15, one line, column 75, 9 characters.

55 The user also identifies data with variable locations. In this particular example, a variable location is specified as follows:

56 Heading 2-line, value="Mail To:"
The identification of Heading 2 as line information means that the system will search for occurrences of the character

13

string "Mail To:" and assign the line number which contains this character string to Heading 2.

The user also identifies data by relative locations. In this example, the user identifies the following relative location:

Mail To-text, Heading 2+1, 3 lines, column 60, 25 characters per line.

The instructions above instruct the system to assign the textual information beginning on one line after Heading 2 and continuing for 3 lines, in column 60, to the Mail To subfile.

As an alternative to inputting the actual line, column, and character numbers, the user can identify desired portions of the document by blocking, or highlighting, the desired portions using the mouse or other input device. In this case, the computer converts the highlighted portions into corresponding line, column, and character numbers.

FIG. 11 lists data corresponding to the hard copy document of FIG. 2 and the associated variable or subfile names.

Next, the user desires to define data relationships in accordance with content instructions. Examples of the type of contents which can be specified by a user are illustrated in FIGS. 12A, 12B, and 12C.

In this particular example, three separate departments of ABC Corporation require information from the XYZ Corporation bill. The first department requires vendor, account number, statement date, payment date, previous balance, new charges, debits, finance charges, payments, and new balance information. The second and third departments require mail to information and previous balance information. Each of these departments have their own application program which utilizes this information.

The user employs content instructions to designate how pieces of information, which have been extracted off of hard copy document 100, are directed to particular departments, that is, particular application programs. FIG. 12A illustrates the contents of the information to be transmitted to the first department. FIG. 12B illustrates the information to be transmitted to the second department. FIG. 12C illustrates the information to be transmitted to the third department. The content instructions, therefore, parse the information shown in FIG. 11 to various application programs, as shown by FIGS. 12A, 12B, and 12C. Content instructions can also be used to identify additional pieces of data which are required for the input files of the particular application programs. In this particular example, the specific application programs from the three departments all require numeric record number information, numeric horizontal position information, numeric vertical position information, and date received information. The horizontal and vertical position information is used by the application program to specify the location of the received information on a spreadsheet application program, in this example. The user may know in advance the content format required by each application program, that is, in this example, the location and type of information specified on the spreadsheet. The user may also employ the split display mode described with reference to FIG. 3B to generate content format instructions.

Using the content instructions, the user establishes the following contents for the input file corresponding to FIG. 12A:

Record number, horizontal position, vertical position, vendor;

Record number, horizontal position, vertical position, account number;

Record number, horizontal position, vertical position, statement date;

Record number, horizontal position, vertical position, date received;

14

Record number, horizontal position, vertical position, payment date;

Record number, horizontal position, vertical position, previous balance;

Record number, horizontal position, vertical position, new charges;

Record number, horizontal position, vertical position, finance charges;

Record number, horizontal position, vertical position, payments;

Record number, horizontal position, vertical position, new balance.

Next, transmission format instructions are employed to

15 define the transmission format of the input file for a specific application program or unit. FIG. 13A illustrates the transmission input file corresponding to FIG. 12A. FIG. 13B illustrates the transmission input file corresponding to FIG. 12B. FIG. 13C illustrates the transmission input file corresponding to FIG. 12C. A comparison of FIGS. 12B and 12C reveals that FIGS. 12B and 12C have the same contents. However, the information illustrated in FIG. 12B is being sent to a different application program than the information in FIG. 12C. These application programs require different transmission input formats as illustrated in FIGS. 13B and 13C. More specifically, the application program that receives the input file illustrated in FIG. 13B uses the greater than sign as a delimiter whereas the application program which receives the transmission input file shown in FIG. 30 13C uses a back-slash as the delimiter.

After the contents and the transmission format for the input file have been defined, and any additional information has been inputted, the input file is assembled and transmitted to the particular application program.

35 FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention. The hard copy document illustrated in FIG. 14 is first scanned and information from the hard copy document is stored into a memory. The interface 200 then identifies 40 portions of the hard copy document corresponding to various variables by recognizing a defined set of symbols. In the FIG. 14 example, triangles delineate the mailing address, circles delineate the statement date, and squares delineate the new charges. Information from these portions of the hard 45 copy document is stored in the corresponding memory locations or subfiles for each variable. The same set of symbols can be used to identify the same information from one document to the next. Thus, even if the physical formats of documents are not fixed from one document to the next, 50 a diversity of hard copy documents can be processed without manually inputting data by recognition of the defined symbols.

55 Examples of readily available application programs are Quicken and Lotus 1,2,3 both of which are widely utilized in the business community. Quicken, for example is an easy-to-utilize program for writing checks and preparing business records. Payee, amount and address information may readily be transmitted from scanner memory 220 and/or main memory 250 to the Quicken application program for 60 check writing functions and ledger keeping purposes. Lotus is a well known spreadsheet program which may process data input into specified cells once this data is placed in conventional Lotus format.

65 Thus, the instant invention provides an integrated and comprehensive system for handling information from a hard copy document, thus permitting a paperless office. In addition, the invention permits data, extracted off of a hard

copy document, to be easily manipulated into various logical and transmission formats required by a particular application unit. The invention also provides a low cost system for inputting information from a wide variety of hard copy documents into a memory.

The foregoing description has been set forth merely to illustrate preferred embodiments of the invention and is not intended to be limiting. Modifications are possible without departing from the scope of the invention.

For example, letters, checks, forms, pictures, reports, music scores, film, and other types of hard copy documents can be processed by the invention for accounts payable/receivable accounting, inventory control, record keeping, budgeting, data base management, music transcription, forms processing, computerized art, survey and questionnaire processing, statistical data analysis, correspondence processing and other applications.

Other automated digitizing units can be used in addition to or as an alternative to use of the scanner 210 as an input unit. Any electrical, magnetic, or optical device which extracts information off of a hard copy document, thereby eliminating the need to manually input significant amounts of information from the hard copy document is suitable for use as an automated digitizing unit. In addition, information can be input by user responses and digital and analog signals generated from various devices, and from computer files from other computer systems. Suitable hardware for inputting data includes a keyboard, a light pen, a mouse, a touch screen, a laser scanner, a microphone, a tablet, a disk drive, a magnetic tape drive, and a modem.

The interface 200 can also output information in forms other than a hard copy of textual or image information. For example, the interface 200 can output system responses, computer files, and digital and analog signals for transmission to other computer systems or to control systems. Suitable hardware for outputting information includes a disk drive, a magnetic tape drive, a cathode ray tube, a plasma screen, a printer, a plotter, a film developer, an amplifier, and a modem.

Since modifications of the described embodiments incorporating the spirit and substance of the invention may occur to persons skilled in the art, the scope of the invention should be limited solely with respect to the appended claims and equivalents.

What is claimed is:

1. A method of inputting information from a diversity of hard copy documents to a computer, comprising at least one of the following steps (a), (b), and (c):

- (a) inputting information from a hard copy document according to an interactive mode which includes the following substeps
 - (a)(1) scanning a hard copy document and displaying a digitized image of at least a portion of the hard copy document on a display;
 - (a)(2) interactively receiving instructions from a user for identifying information on said digitized image and converting at least said identified information to converted text data corresponding to at least one field of information required by an application program; and
 - (a)(3) transferring said converted text data to said field of information;
- (b) creating a template to enable transfer of converted text data corresponding to at least one field of information required by an application program according to an interactive template creation mode which includes the following substeps

(b)(1) scanning a hard copy document and displaying an image of at least a portion of the hard copy document on a display;

(b)(2) interactively receiving instructions from a user to create a template which identifies at least one location on said image as containing information to be converted to text data corresponding to a field of information required by an application program; and

(b)(3) storing said template;

(c) inputting information from a hard copy document according to an automatic mode which includes the following substeps

(c)(1) scanning a hard copy document to generate document information; and

(c)(2) matching at least a portion of said document information with a template created to transfer converted text data to at least one field of information required by an application program.

2. A method as set forth in claim 1, further comprising the

steps of:

defining a set of symbols which designate fields of information required by an application program; and detecting the presence of a particular one of said defined set of symbols on a hard copy document and extracting a field of information required by an application program based on said detecting.

3. A method as set forth in claim 1, further comprising the

step of:

displaying a split-screen image showing at least a portion of said hard copy document and at least one field of information required by an application program during the interactive mode.

4. A method as set forth in claim 1, further comprising the

step of:

identifying said located information in step (a)(2) by blocking portions of said image on said display using an input device.

5. A method as set forth in claim 3, further comprising the

step of:

identifying said located information in step (a)(2) by blocking portions of said image on said display using an input device.

6. A method as set forth in claim 1, further comprising the

step of:

displaying a split-screen image showing at least a portion of said hard copy document and at least one field of information required by an application program during the interactive template creation mode.

7. A method as set forth in claim 1, further comprising the

step of:

creating a template in step (b)(2) by blocking portions of said image on said display using an input device.

8. A method as set forth in claim 6, further comprising the

step of:

creating a template in step (b)(2) by blocking portions of said image on said display using an input device.

9. A method as set forth in claim 1, further comprising the

step of:

storing a digitized image of at least a portion of a hard copy document in an image memory.

10. A method as set forth in claim 1, further comprising the

step of:

storing text data from the hard copy document in a text memory.

11. A method as set forth in claim 1, further comprising the

step of:

storing a digitized image of at least a portion of a hard copy document in an image memory; and
storing text data from the hard copy document in a text memory.

12. A method as set forth in claim 1, further comprising the step of:

repeating substeps (a)(2) and (a)(3) at least once for other identified information and another field of information.

13. A method as set forth in claim 12, further comprising the steps of repeating substeps (a)(2) and (a)(3) for another 10 different application program.

14. A method as set forth in claim 1, further comprising the steps of:

repeating substeps (a)(1), (a)(2) and (a)(3) for another 15 different application program; and

creating another template for another different application program.

15. A method as set forth in claim 1, further comprising the step of:

repeating substeps (c)(1) and (c)(2) for another different hard copy document.

16. A method as set forth in claim 1 wherein in step (b)(2) said template identifies a plurality of locations on said image on said display as containing a field of information required by an application program.

17. A method as set forth in claim 16, further comprising the step of creating another template for another different application program.

18. A method as set forth in claim 1, further comprising at least one of the steps of:

repeating substeps (a)(2) and (a)(3) for another different application program; and

creating another template for another different application program.

19. A method as set forth in claim 1, further comprising the steps of repeating substeps (a)(1), (a)(2) and (a)(3) for another different application program.

20. A method of inputting information from a diversity of hard copy documents to a computer, comprising at least two of the following steps (a), (b), and (c):

(a) inputting information from a hard copy document according to an interactive mode which includes the following substeps

(a)(1) scanning a hard copy document and displaying a digitized image of at least a portion of the hard copy 45 document on a display;

(a)(2) interactively receiving instructions from a user for identifying information on said digitized image and converting at least said identified information to converted text data corresponding to at least one field of information required by an application program; and

(a)(3) transferring said converted text data to said field of information;

(b) creating a template to enable transfer of converted text data corresponding to at least one field of information required by an application program according to an interactive template creation mode which includes the following substeps

(b)(1) scanning a hard copy document and displaying 60 an image of at least a portion of the hard copy document on a display;

(b)(2) interactively receiving instructions from a user to create a template which identifies at least one location on said image as containing information to be converted to text data corresponding to a field of information required by an application program; and

(b)(3) storing said template;

(c) inputting information from a hard copy document according to an automatic mode which includes the following substeps

(c)(1) scanning a hard copy document to generate document information; and

(c)(2) matching at least a portion of said document information with a template created to transfer converted text data to at least one field of information required by an application program.

21. A method as set forth in claim 20, further comprising the steps of:

defining a set of symbols which designate fields of information required by an application program; and detecting the presence of a particular one of said defined set of symbols on a hard copy document and extracting a field of information required by an application program based on said detecting.

22. A method as set forth in claim 20, further comprising the step of:

displaying a split-screen image showing at least a portion of said hard copy document and at least one field of information required by an application program during the interactive mode.

23. A method as set forth in claim 20, further comprising the step of:

identifying said information in step (a)(2) by blocking portions of said image on said display using an input device.

24. A method as set forth in claim 22, further comprising the step of:

identifying said information in step (a)(2) by blocking portions of said image on said display using an input device.

25. A method as set forth in claim 20, further comprising the step of:

displaying a split-screen image showing at least a portion of said hard copy document and at least one field of information required by an application program during the interactive template creation mode.

26. A method as set forth in claim 20, further comprising the step of:

creating a template in step (b)(2) by blocking portions of said image on said display using an input device.

27. A method as set forth in claim 25, further comprising the step of:

creating a template in step (b)(2) by blocking portions of said image on said display using an input device.

28. A method as set forth in claim 20, further comprising the step of:

storing a digitized image of at least a portion of a hard copy document in an image memory.

29. A method as set forth in claim 20, further comprising the step of:

storing text data from the hard copy document in a text memory.

30. A method as set forth in claim 20, further comprising the step of:

storing a digitized image of at least a portion of a hard copy document in an image memory; and

storing text data from the hard copy document in a text memory.

31. A method as set forth in claim 20, further comprising the step of:

repeating substeps (a)(2) and (a)(3) at least once for other identified information and another field of information.

32. A method as set forth in claim 31, further comprising the steps of repeating substeps (a)(2) and (a)(3) for another different application program.

33. A method as set forth in claim 20, further comprising the steps of:

repeating substeps (a)(1), (a)(2) and (a)(3) for another different application program; and
creating another template for another different application program.

34. A method as set forth in claim 20, further comprising the step of:

repeating substeps (c)(1) and (c)(2) for another different hard copy document.

35. A method as set forth in claim 20, wherein in step (b)(2) said template identifies a plurality of locations on said image on said display as containing a field of information required by an application program.

36. A method as set forth in claim 20, further comprising the step of creating another template for another different application program.

37. A method as set forth in claim 20, further comprising the steps of:

repeating substeps (a)(2) and (a)(3) for another different application program; and
creating another template for another different application program.

38. A method as set forth in claim 20, further comprising the steps of repeating substeps (a)(1), (a)(2) and (a)(3) for another different application program.

39. A method as set forth in claim 20, further comprising at least two of the steps of:

(a) repeating substeps (a)(1), (a)(2) and (a)(3) at least once for another different hard copy document;
(b) repeating substeps (b)(1), (b)(2) and (b)(3) at least once for another different hard copy document; and
(c) repeating substeps (c)(1) and (c)(2) for another different hard copy document.

40. A method as recited in claim 39 wherein said at least two step are steps (a) and (b).

41. A method as recited in claim 39 wherein said at least two step are steps (a) and (c).

42. A method as recited in claim 39 wherein said at least two step are steps (b) and (c).

43. A method of inputting information from a diversity of hard copy documents to a computer, comprising each of the following steps (a), (b), and (c):

(a) inputting information from a hard copy document according to an interactive mode which includes the following substeps

(a)(1) scanning a hard copy document and displaying a digitized image of at least a portion of the hard copy document on a display;

(a)(2) interactively receiving instructions from a user for identifying information on said digitized image and converting at least said identified information to converted text data corresponding to at least one field of information required by an application program; and

(a)(3) transferring said converted text data to said field of information;

(b) creating a template to enable transfer of converted text data corresponding to at least one field of information required by an application program according to an interactive template creation mode which includes the following substeps

(b)(1) scanning a hard copy document and displaying an image of at least a portion of the hard copy document on a display;

(b)(2) interactively receiving instructions from a user to create a template which identifies at least one location on said image as containing information to be converted to text data corresponding to a field of information required by an application program; and
(b)(3) storing said template;

(c) inputting information from a hard copy document according to an automatic mode which includes the following substeps

(c)(1) scanning a hard copy document to generate document information; and
(c)(2) matching at least a portion of said document information with a template created to transfer converted text data to at least one field of information required by an application program.

44. A method as set forth in claim 43, further comprising the steps of:

defining a set of symbols which designate fields of information required by an application program; and detecting the presence of a particular one of said defined set of symbols on a hard copy document and extracting a field of information required by an application program based on said detecting.

45. A method as set forth in claim 43, further comprising the step of:

displaying a split-screen image showing at least a portion of said hard copy document and at least one field of information required by an application program during the interactive mode.

46. A method as set forth in claim 43, further comprising the step of:

identifying said information in step (a)(2) by blocking portions of said image on said display using an input device.

47. A method as set forth in claim 45, further comprising the step of:

identifying said information in step (a)(2) by blocking portions of said image on said display using an input device.

48. A method as set forth in claim 43, further comprising the step of:

displaying a split-screen image showing at least a portion of said hard copy document and at least one field of information required by an application program during the interactive template creation mode.

49. A method as set forth in claim 43, further comprising the step of:

creating a template in step (b)(2) by blocking portions of said image on said display using an input device.

50. A method as set forth in claim 48, further comprising the step of:

creating a template in step (b)(2) by blocking portions of said image on said display using an input device.

51. A method as set forth in claim 43, further comprising the step of:

storing a digitized image of at least a portion of a hard copy document in an image memory.

52. A method as set forth in claim 43, further comprising the step of:

storing text data from the hard copy document in a text memory.

53. A method as set forth in claim 43, further comprising the step of:

21

storing a digitized image of at least a portion of a hard copy document in an image memory; and

storing text data from the hard copy document in a text memory.

54. A method as set forth in claim **43**, further comprising the step of:

repeating substeps (a)(2) and (a)(3) at least once for other identified information and another field of information.

55. A method as set forth in claim **54**, further comprising the steps of repeating substeps (a)(2) and (a)(3) for another different application program.

56. A method as set forth in claim **43**, further comprising the step of:

repeating substeps (a)(1), (a)(2) and (a)(3) for another different application program; and

creating another template for another different application program.

57. A method as set forth in claim **43**, further comprising the step of:

repeating substeps (c)(1) and (c)(2) for another different hard copy document.

58. A method as set forth in claim **43**, wherein in step (b)(2) said template identifies a plurality of locations on said image on said display as containing a field of information required by an application program.

59. A method as set forth in claim **43**, further comprising the step of creating another template for another different application program.

60. A method as set forth in claim **43**, further comprising the steps of:

repeating substeps (a)(2) and (a)(3) for another different application program; and

22

creating another template for another different application program.

61. A method as set forth in claim **1**, **20** or **43** further comprising at least one of the steps of:

(a) repeating substeps (a)(1), (a)(2) and (a)(3) at least once for another different hard copy document;

(b) repeating substeps (b)(1), (b)(2) and (b)(3) at least once for another different hard copy document; and

(c) repeating substeps (c)(1) and (c)(2) for another different hard copy document.

62. A method as set forth in claim **61** wherein said at least one step is step (a).

63. A method as set forth in claim **61** wherein said at least one step is step (b).

64. A method as set forth in claim **61** wherein said at least one step is step (c).

65. A method as set forth in claim **54**, further comprising the steps of repeating substeps (a)(1), (a)(2) and (a)(3) for another different application program.

66. A method as set forth in claim **43**, further comprising the steps of:

repeating substeps (a)(1), (a)(2) and (a)(3) at least once for another different hard copy document;

repeating substeps (b)(1), (b)(2) and (b)(3) at least once for another different hard copy document; and

repeating substeps (c)(1) and (c)(2) for another different hard copy document.

* * * * *

United States Patent [19]

Lech et al.

[11] Patent Number: 6,094,505

[45] Date of Patent: Jul. 25, 2000

[54] INFORMATION PROCESSING
METHODOLOGY[75] Inventors: **Robert Lech**, Jackson; **Mitchell A. Medina**, Essex Fells; **Catherine B. Elias**, Plainsboro, all of N.J.[73] Assignee: **Millennium L.P.**, Grand Cayman, Cayman Islands

[21] Appl. No.: 09/044,159

[22] Filed: Mar. 19, 1998

Related U.S. Application Data

[60] Continuation of application No. 08/487,150, Jun. 7, 1995, Pat. No. 5,768,416, which is a division of application No. 08/348,224, Nov. 28, 1994, Pat. No. 5,625,465, which is a continuation of application No. 08/143,135, Oct. 29, 1993, Pat. No. 5,369,508, which is a continuation of application No. 07/672,865, Mar. 20, 1991, Pat. No. 5,258,855.

[51] Int. Cl.⁷ G06F 15/40

[52] U.S. Cl. 382/180; 382/282

[58] Field of Search 382/287, 175, 382/306, 177, 180, 282, 317; 395/148, 149

[56] References Cited

U.S. PATENT DOCUMENTS

4,034,343	7/1977	Wilmer	340/146.3 MA
4,387,964	6/1983	Arrazika et al.	350/331 R
4,593,367	6/1986	Slack et al.	364/513
4,667,248	5/1987	Kanno	358/280
4,760,606	7/1988	Lesnick et al.	382/48
4,802,104	1/1989	Ogiso	364/518
4,931,957	6/1990	Takagi et al.	364/521
4,933,979	6/1990	Suzuki et al.	381/61
5,031,121	7/1991	Iwai et al.	364/523
5,034,990	7/1991	Klees	382/22
5,095,445	3/1992	Sekiguchi	364/514
5,140,650	8/1992	Casey et al.	382/61
5,153,927	10/1992	Yamanari	382/61

5,228,100	7/1993	Takeda et al.	382/61
5,258,855	11/1993	Lech et al.	358/462
5,367,619	11/1994	Dipaolo	395/149
5,404,294	4/1995	Karnik	364/419.1
5,416,849	5/1995	Huang	382/173
5,448,738	9/1995	Good et al.	395/700
5,452,379	9/1995	Poor	382/317
5,506,697	4/1996	Li et al.	358/448

OTHER PUBLICATIONS

TopScan Professional User's Guide, Calera Recognition Systems, pp. v–vii, xi–xiii, and 1–6, 1989.

PC Magazine, vol. 5, Number 16, Sep. 30, 1986.

Que's Computer User's Dictionary, 2nd Ed., Bryan Pfaffenberger (author); 1991; p. 144.

“Kurzweil 5200 Intelligent Scanning System”, Xerox Imaging Systems, Inc., 1990.

Primary Examiner—Thomas D. Lee

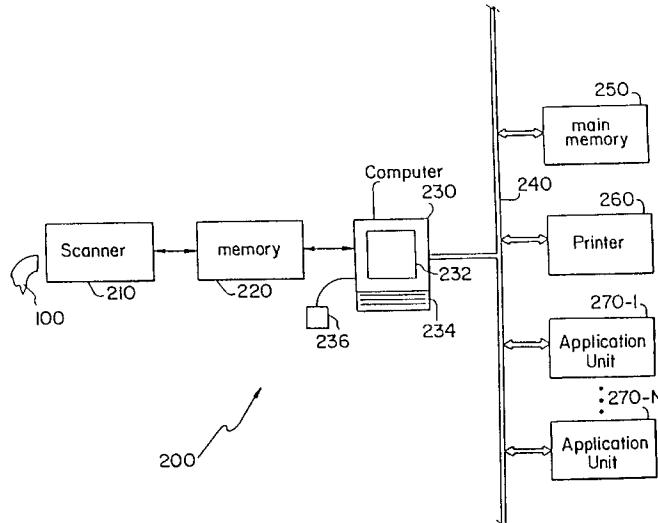
Assistant Examiner—Stephen Brinich

Attorney, Agent, or Firm—Foley & Lardner

[57] ABSTRACT

An information processing methodology gives rise to an application program interface which includes an automated digitizing unit, such as a scanner, which inputs information from a diversity of hard copy documents and stores information from the hard copy documents into a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which designate portions of the stored document information required by a particular application program. The selected stored document information is then placed into the transmission format required by a particular application program in accordance with transmission format instructions. After the information has been transmission formatted, the information is transmitted to the application program. In one operational mode, the interface interactively prompts the user to identify, on a display, portions of the hard copy documents containing information used in application programs or for storage.

20 Claims, 15 Drawing Sheets



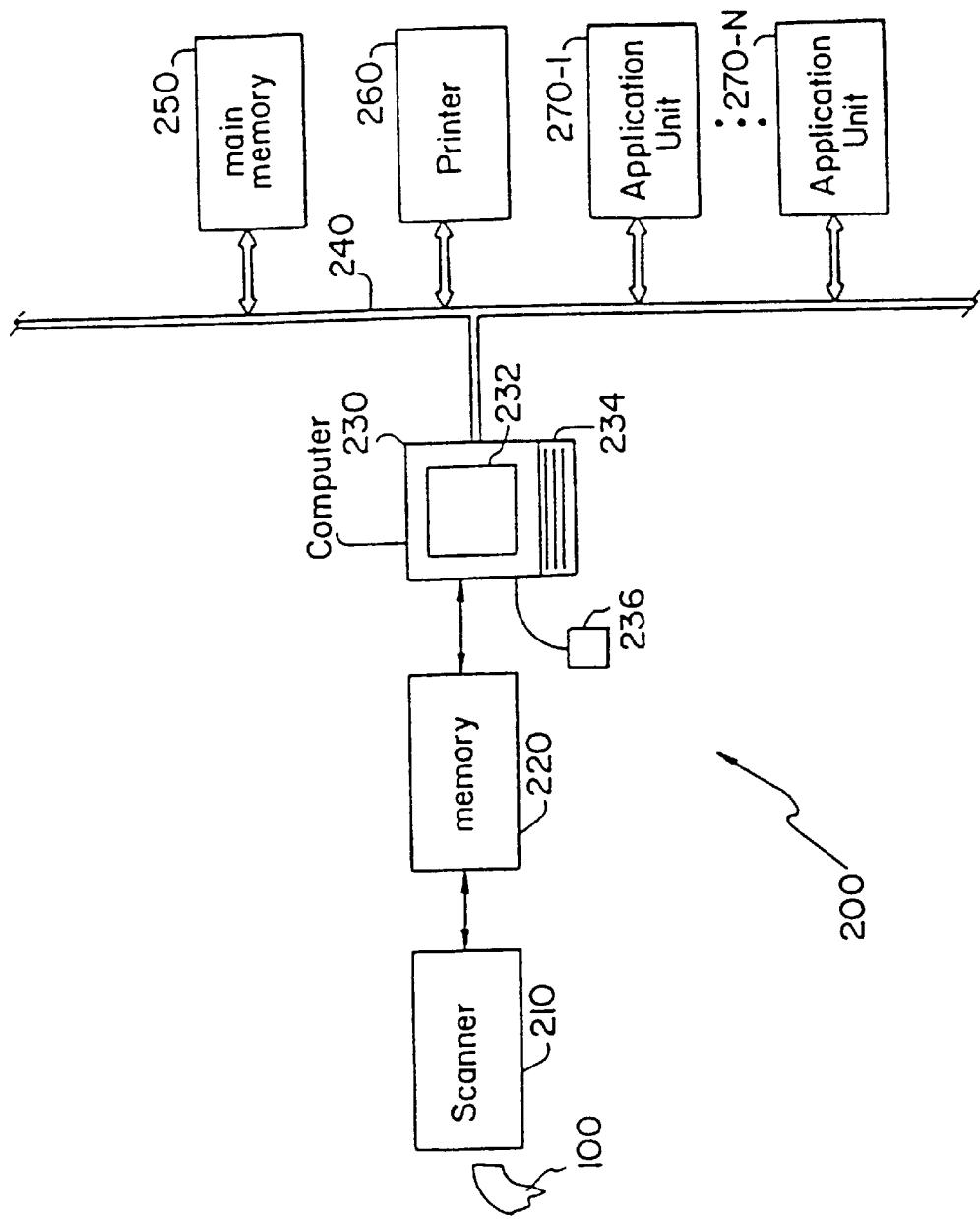
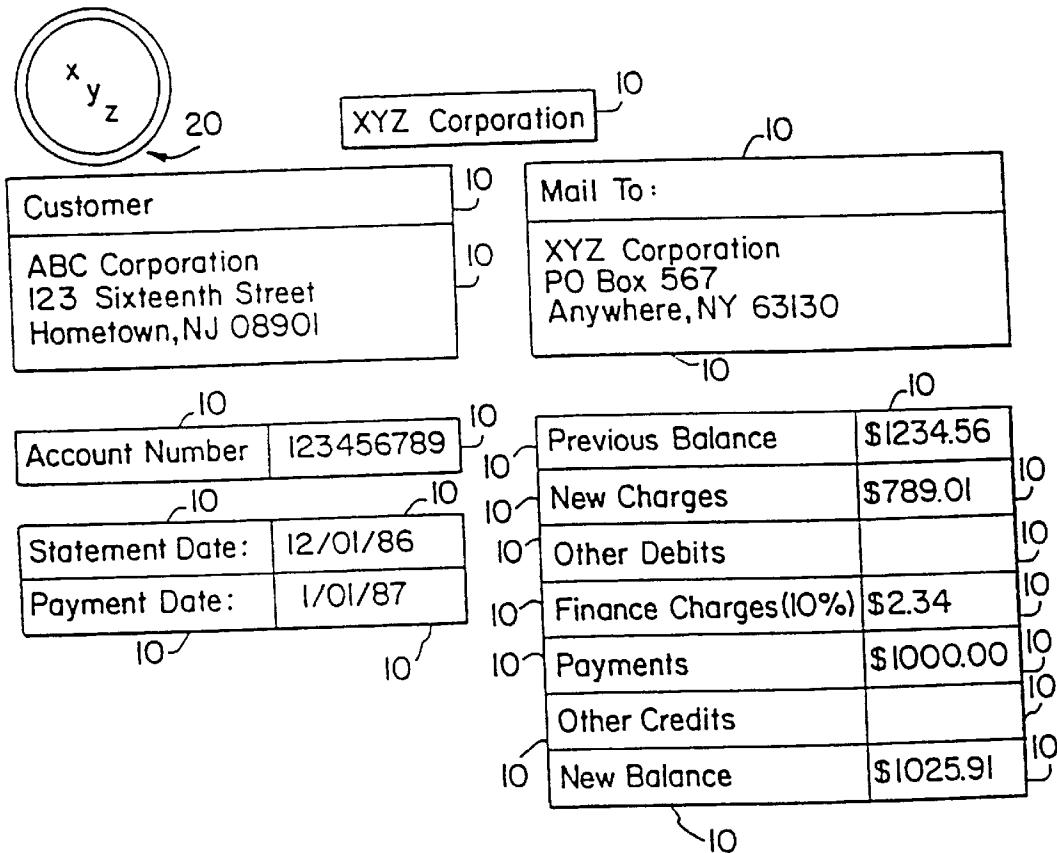


FIG. I



Winter Sale In Effect Throughout January!

FIG. 2

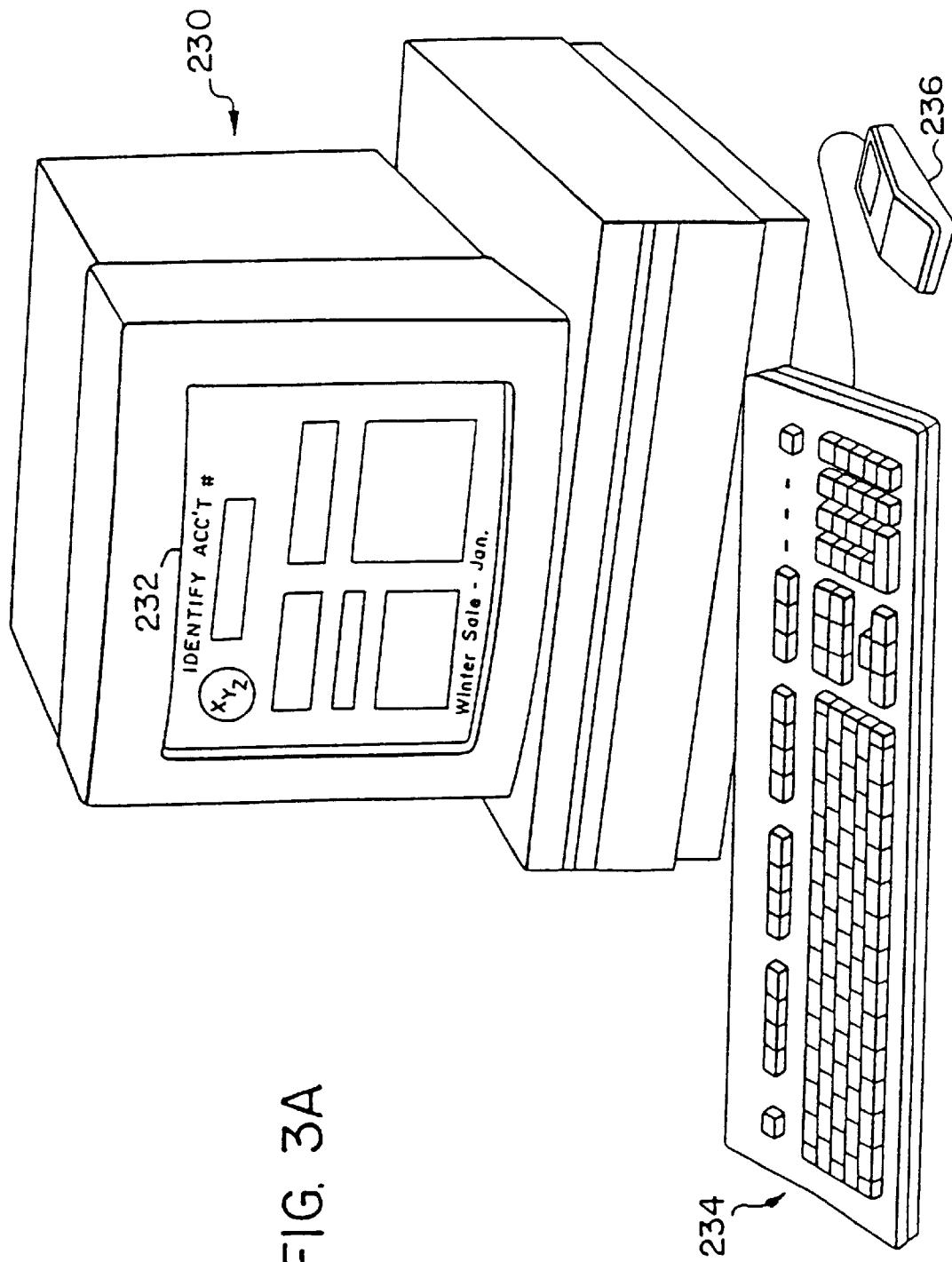


FIG. 3A

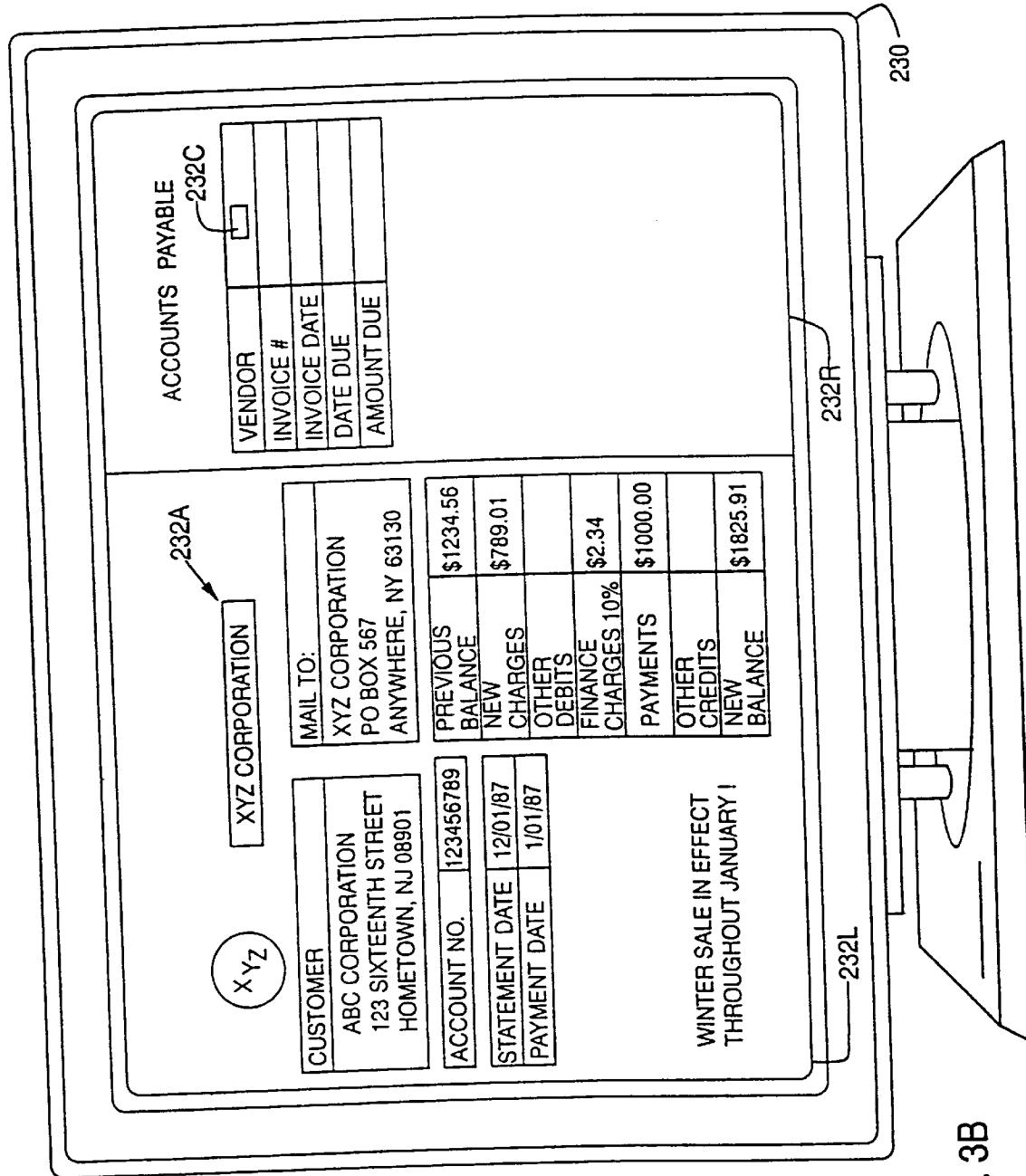


FIG. 3B

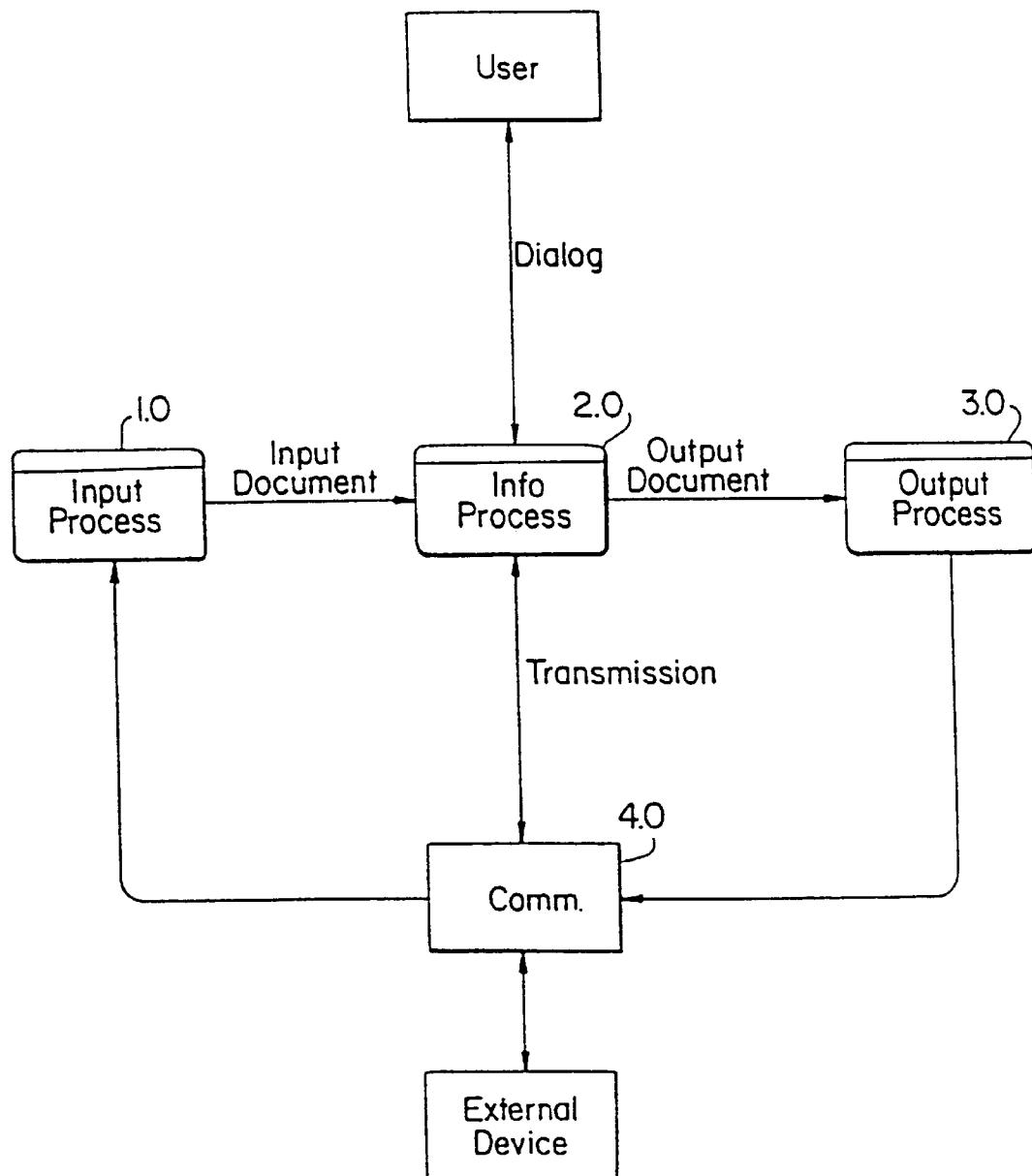
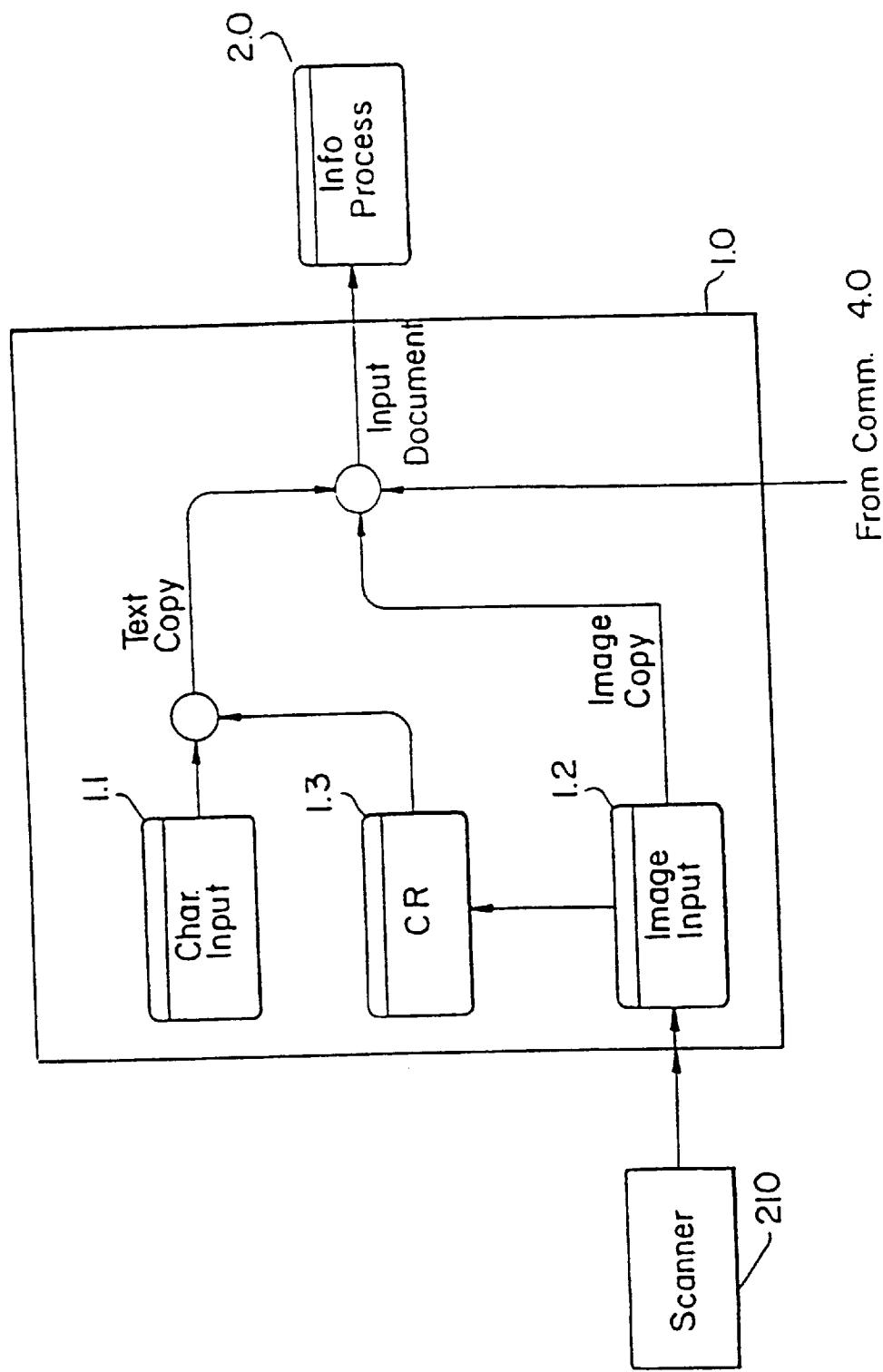


FIG. 4



5
FIG.

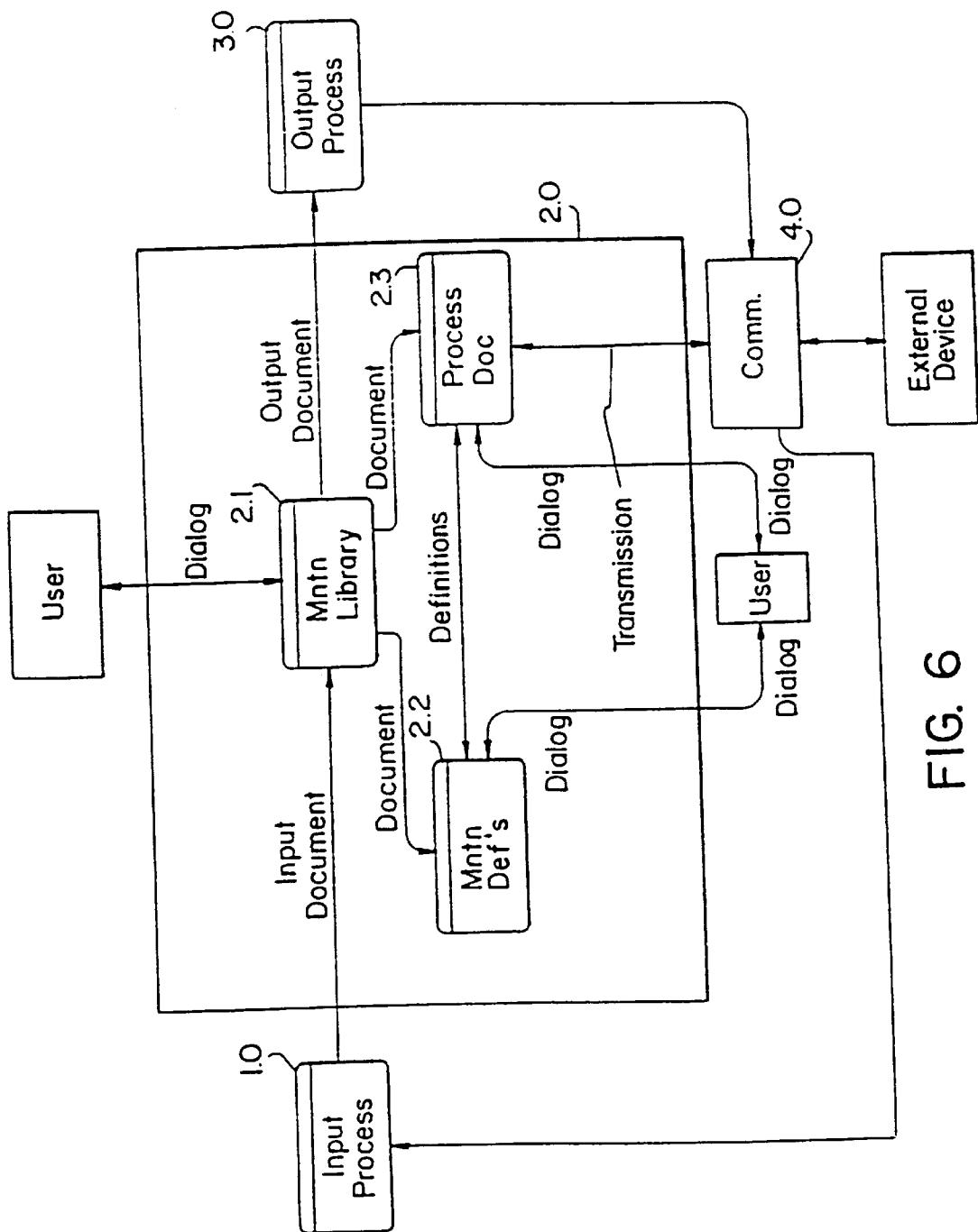


FIG. 6

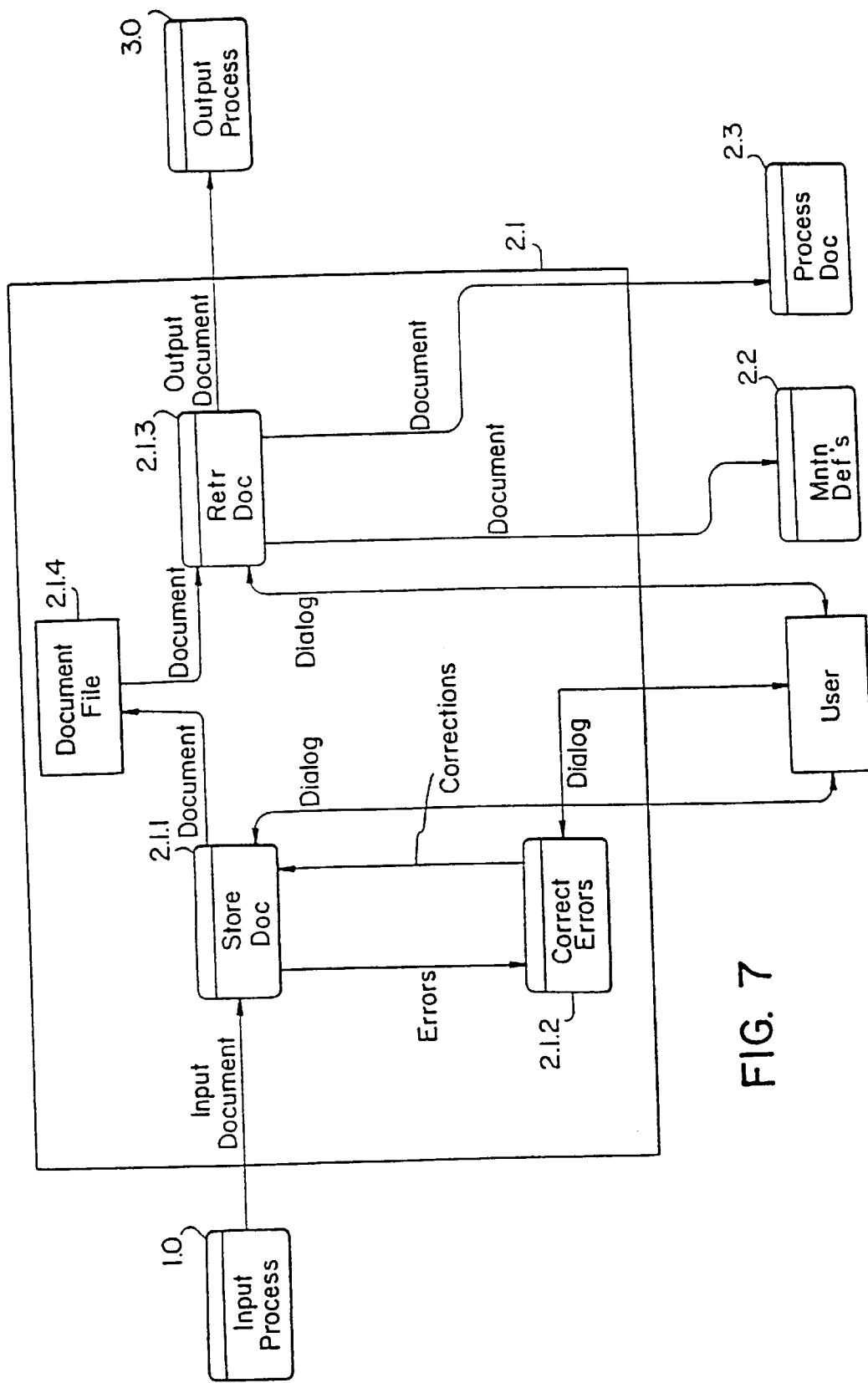


FIG. 7

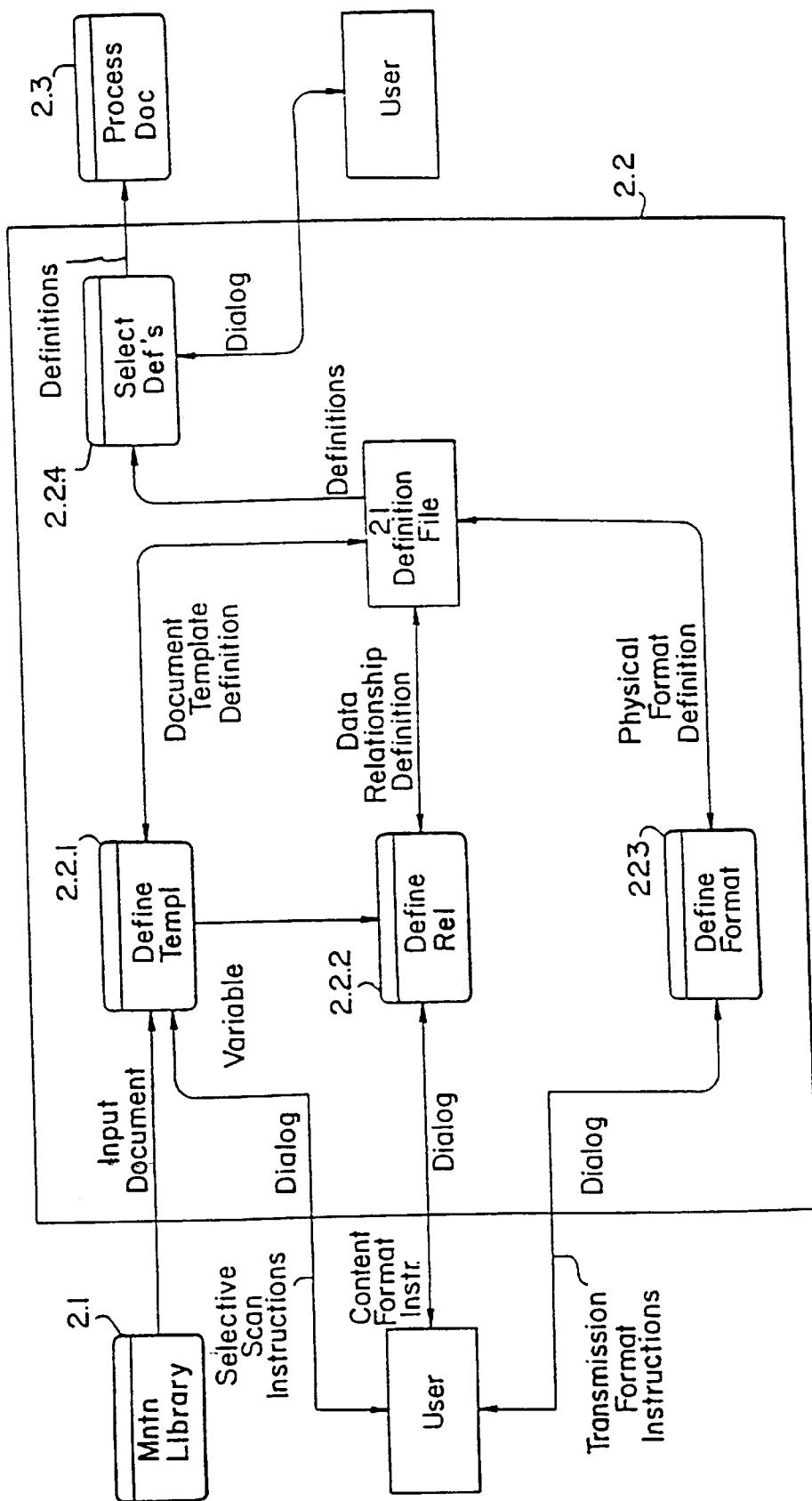


FIG. 8

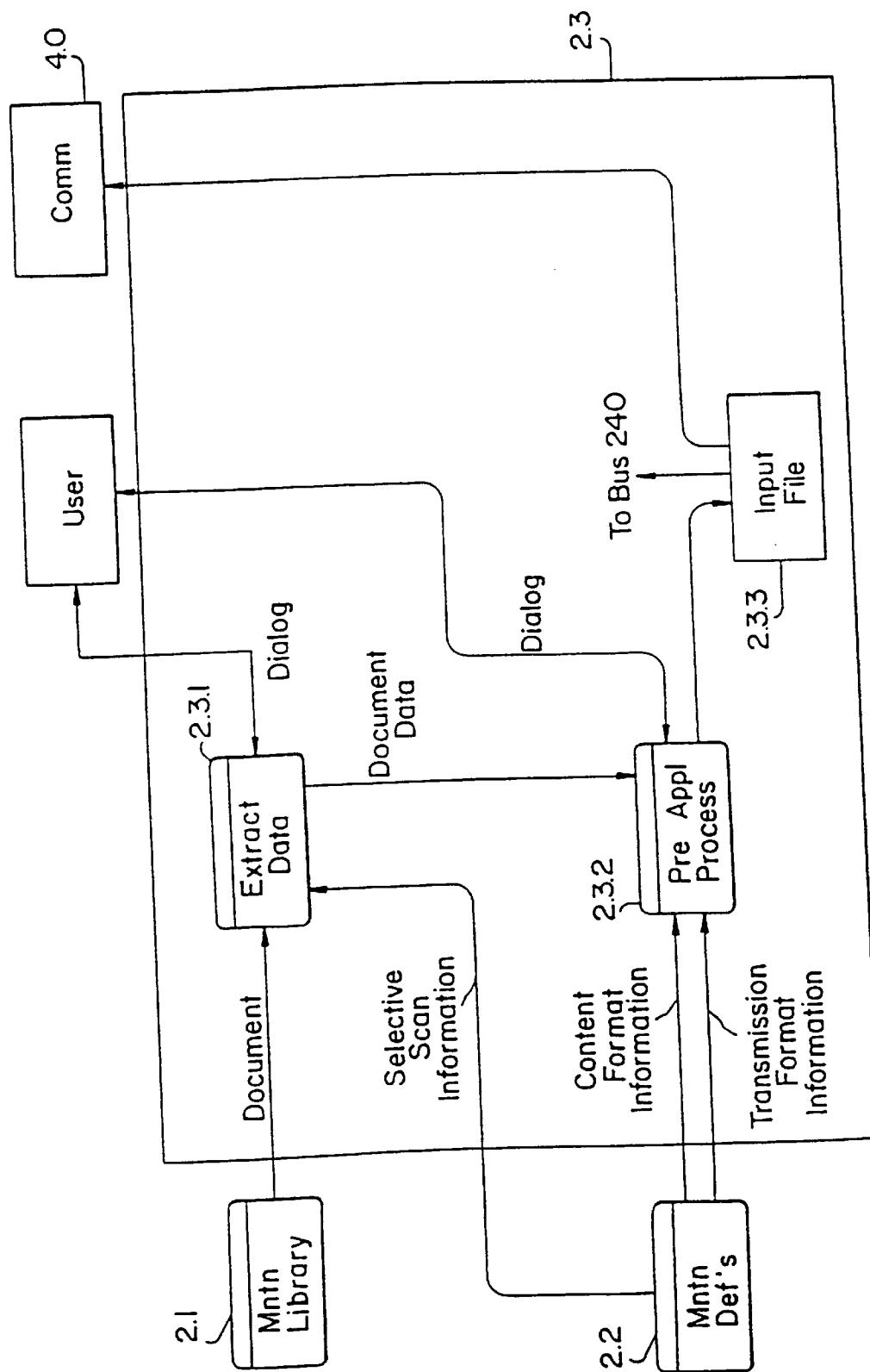


FIG. 9

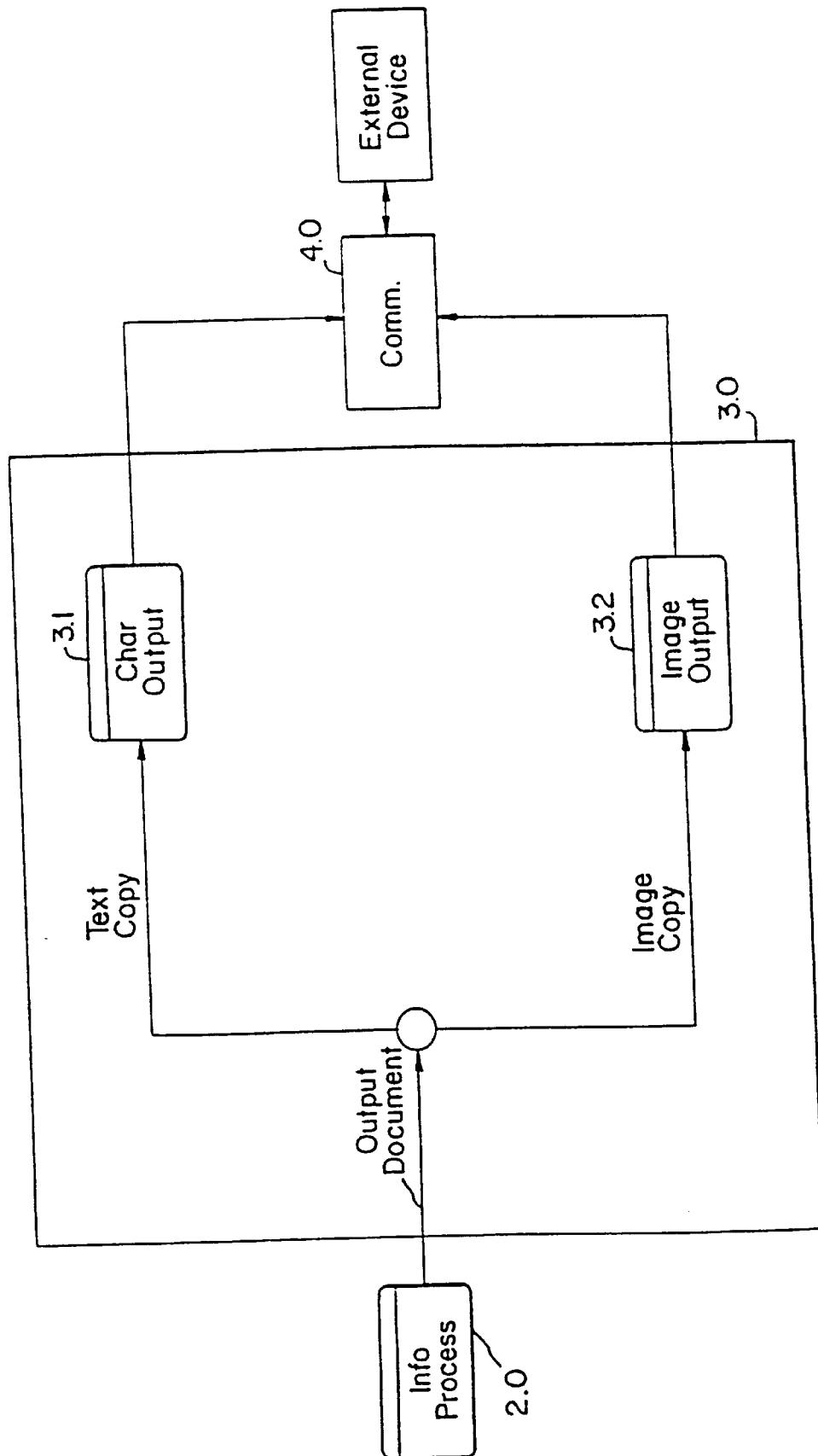


FIG. 10

FIG. 11

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Heading 2	2
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
Other Credits	
New Balance	\$1025.91

FIG. 12A

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
New Balance	\$1025.91

FIG. 12B

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

FIG. 12C

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

FIG.13A

```
> 1 > 1 > 1 > "XYZ Corporation"  
> 2 > 2 > 25 >+ 123456789 >  
> 3 > 2 > 1 > D12/01/86 >  
> 4 > 2 > 11 > D12/15/86 >  
> 5 > 2 > 21 > D01/01/87 >  
> 6 > 10 > 25 > $1234.56 >  
> 7 > 11 > 25 > $789.01 >  
> 8 > 13 > 25 > $2.34 >  
> 9 > 14 > 25 > $1000.00 >  
> 10 > 16 > 25 > $1025.91 >
```

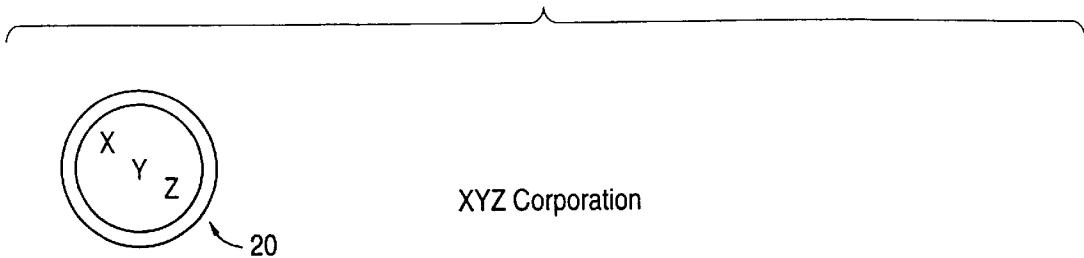
FIG.13B

```
> 1 > 1 > 1 > $1234.56 >  
> 2 > 2 > 1 > "XYZ Corporation"  
> 3 > 3 > 1 > "PO Box 567"  
> 4 > 4 > 1 > "Anywhere, NY 63130"
```

FIG.13C

```
/1/1/1/$1234.56//  
/2/2/1/*XYZ Corporation*  
/3/3/1/*PO Box 567*  
/4/4/1/*Anywhere, NY 63130*
```

FIG. 14



Customer

ABC Corporation
123 Sixteenth Street
Hometown, NJ 88981

Mail To:

△ XYZ Corporation
PO Box 567
Anywhere, NY 63130 △

Account Number 123456789

Previous Balance \$1234.56

Statement Date: 12/01/86 New Charges \$789.01

Payment Date: 1/01/87

Other Debits

Finance Charges (10%) \$2.34

Payments \$1000.00

WINTER SALE IN EFFECT THROUGHOUT JANUARY: Other Credits

New Balance \$1025.91

1

INFORMATION PROCESSING
METHODOLOGY

This application is a continuation, of application Ser. No. 08/487,150, filed Jun. 7, 1995, U.S. Pat. No. 5,768,416 which is a divisional of Ser. No. 08/348,224, filed Nov. 28, 1994 (now U.S. Pat. No. 5,625,465), which is a continuation Ser. No. 08/143,135, filed Oct. 29, 1993 (now U.S. Pat. No. 5,369,508), which is a continuation of Ser. No. 07/672,865, filed Mar. 20, 1991 (now U.S. Pat. No. 5,258,855).

BACKGROUND OF THE INVENTION

The invention is directed to a system for efficiently processing information originating from hard copy documents. More specifically, the invention is directed to a hard copy document to application program interface which minimizes the need to manually process hard copy documents.

In the past, information contained on hard copy documents was manually entered into a computer via the input controller of a particular computer. The original document was then filed away for future reference. Automatic input of data was limited to the input of Magnetic Ink Character Recognition (MICR) data and to Optical Character Recognition (OCR) data. This fixed-position data was forwarded directly to a dedicated computer application specifically designed to accommodate the input format. In more recent years, typewritten text has been mechanically inputted into a computer via a text file. Examples of this latter type of system are word processors and photo-typesetters.

These conventional systems have limitations which decrease the efficiency of processing information from a hard copy document. For example, the systems discussed above are limited in their application to MICR, OCR, or typewritten data. Parsing and processing data is limited to the particular requirements of the particular computer application which requires the input data. In addition, in these conventional systems, the actual hard copy document must be retained for future reference at great expense.

In a sophisticated computer network, different users may require different portions of the information contained on a hard copy document. For example, if the hard copy document is an invoice returned with payment of a bill, the accounting department may need all of the monetary information contained on the bill while the mailroom may need only customer address information, to update a customer's address. Therefore, there is a need for a system in which specific information from a hard copy document can be selectively distributed to various users.

Another problem with conventional systems is that users, even within the same company, may require that the information extracted from a hard copy document be transmitted to a particular application program in a specific transmission format. For example, one department in a company may use a particular application program which must receive information using a particular character as a delimiter and other departments may require the information in a different format using different delimiters.

Another problem, particularly for small businesses, is that current systems can not efficiently accommodate the inputting of information from a diversity of hard copy documents. A large business which receives many forms in the same format can afford a system which inputs a high volume of information in that format into memory. For example, it is cost-effective for a bank which processes hundreds of thousands of checks a month to buy a dedicated machine which

2

can read information off of checks having a rigidly defined, or fixed, format. However, as the diversity of forms received by a business increases relative to the number of forms that must be processed, it becomes less cost-effective to design a dedicated machine for processing each type of form format. This problem is particularly significant in small businesses which may, for example, receive fifty invoices a month, all in different, non-fixed, formats. It is frequently not cost-effective for a small business to design dedicated systems for inputting information in each of these various formats. This leaves a small business with no other practical alternative than to manually input the information off of each invoice each month.

SUMMARY OF THE INVENTION

It is an object of the invention, therefore, to provide an application program interface which allows a user to select specific portions of information extracted from a diversity of hard copy documents and allows the user to direct portions of this information to several different users in accordance with the needs of the particular user.

It is also an object of the invention to provide a cost-effective system for inputting hard copy documents which can accommodate hard copy documents in a diversity of formats.

It is another object of the invention to provide an application program interface which allows a user to put information, which is to be transmitted, into a particular transmission format, based upon the needs of the receiver of the information.

It is a further object of the invention to provide an application program interface which will allow the extraction, selection, formatting, routing, and storage of information from a hard copy document in a comprehensive manner such that the hard copy document itself need not be retained.

It is another object of the invention to provide a system which reduces the amount of manual labor required to process information originating from a hard copy document.

A further object of the invention is to reduce the time required to process information originating from a hard copy document so that a higher volume of transactions involving hard copy documents can be processed.

The invention provides an application program interface which inputs a diversity of hard copy documents using an automated digitizing unit and which stores information from the hard copy documents in a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which define portions of the stored document information required by a particular application unit. Selected stored document information is then formatted into the transmission format used by the particular application program based on transmission format instructions. The transmission formatted selected stored document information is then transmitted to the particular application program. The hard copy documents may contain textual information or image information or both.

The interface operates in three different modes.

In a first mode, the interface extracts all of the information from hard copy documents and stores this information in memory. Parsing of various portions of the extracted information is performed in accordance with content instructions.

In a second mode, the user operates interactively with the interface by use of a display and an input device, such as a

mouse. In this second mode, a hard copy document is inputted and displayed on the display. The interface then prompts the user to identify the location of various information. For example, the interface can ask the user to identify the location of address information on the hard copy document. In response, the user positions the mouse to identify address information using a cursor. The identified information is then stored as address information in memory. Subsequently, the interface again prompts the user to identify other pieces of information, which are then stored in the appropriate locations in memory. This process proceeds until all of the information which is desired to be extracted off of the hard copy document is stored in memory.

In a third mode of operation, selected portions of information are extracted off of hard copy documents in accordance with predetermined location information which has been specified by the user. For example, the user can define a template which specifies the location of information on hard copy documents. Templates can be formed in conjunction with second mode operation. Alternatively, the user can instruct the interface to search hard copy documents for a particular character or symbol, located on the hard copy documents. The information desired to be extracted off of the hard copy documents is specified relative to the location of this character or symbol.

The interface can also prompt or receive from an applications program or another information processing system, required information, content instructions, and format instructions.

Other objects, features, and advantages of the invention will be apparent from the following detailed description of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be described in further detail below with reference to the accompanying drawings, in which:

FIG. 1 illustrates hardware for implementing a preferred embodiment of the instant invention;

FIG. 2 illustrates an example of a hard copy document containing information to be processed by the instant invention;

FIGS. 3A and 3B are enlarged views of the computer of FIG. 1 used to explain how the invention interactively prompts a user to identify information;

FIG. 4 is an overall data flow diagram for the FIG. 1 preferred embodiment;

FIG. 5 is a detailed input data flow diagram for the FIG. 1 preferred embodiment;

FIG. 6 is a detailed information processing data flow diagram for the FIG. 1 preferred embodiment;

FIG. 7 is a more detailed information processing data flow diagram for the maintain library module of FIG. 6;

FIG. 8 is a more detailed information processing data flow diagram for the maintain definitions module of FIG. 6;

FIG. 9 is a more detailed information processing data flow diagram for the process document module of FIG. 6;

FIG. 10 is a detailed output data flow diagram for the FIG. 1 preferred embodiment;

FIG. 11 lists data corresponding to the hard copy document of FIG. 2;

FIGS. 12A, 12B, and 12C illustrate examples of data which can be selected from the extracted data of FIG. 11 in accordance with content instructions;

FIGS. 13A, 13B, and 13C illustrate examples of the data of FIGS. 12A, 12B, and 12C formatted in accordance with various transmission format instructions to form input files; and

FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Hardware

The invention provides an interface between information originating from a hard copy document and a computer application unit which uses the information. The computer application unit can be a particular computer application program or a device which is controlled in accordance with instructions or information from the hard copy document.

10 The invention also allows storing a copy of the hard copy document in a memory and retrieving the copy of the hard copy document. By providing a comprehensive and integrated system which can accommodate almost all of the possible uses of information contained on a hard copy document, the instant invention allows for a paperless office.

15 The invention includes hardware and software necessary to extract, retrieve, and process information from the hard copy document. A copy of the actual image of the hard copy document is stored in memory. Textual information extracted from the hard copy document is also stored in memory. Textual information is information, such as alphanumeric characters, which is recognized on the hard copy document and which is stored in a form which corresponds to the particular recognized character. For example, the 20 extracted characters can be stored in the ASCII format in an electronic memory.

25 The user can have all of the information extracted from the hard copy document and stored in memory. Alternatively, the interface can interactively prompt the user to identify specific pieces of information for storage. The interface can also extract specific pieces of information using a predefined template. The interface can also prompt or receive from another information processing system or an 30 applications program desired information, content instructions, and format instructions.

35 The instant invention also provides for parsing information extracted from the hard copy document and for directing this parsed information to specific users or application programs as an input file.

40 The invention also permits the user to define the transmission format of the input file for a particular computer application unit.

45 FIG. 1 illustrates hardware for implementing a preferred embodiment of a hard copy document to application program interface according to the instant invention. The interface 200 processes information extracted off of hard copy document 100 and provides information to application units 270 in a form required by each particular application unit. 50 The interface extracts information off of a hard copy document 100 utilizing a scanner 210. The scanner 210 can be any type of scanner which extracts information off of hard copy documents, for example, an Optical Reader.

55 The scanned information is stored in a scanner memory 220 or in main memory 250, as will be described in greater detail below. If main memory 250 or another memory is available to store the scanned information, then scanner memory 220 can be omitted.

60 The information from scanner memory 220 or main memory 250 is transmitted to computer 230. In the preferred embodiment, computer 230 includes a display 232, a keyboard 234, and a mouse 236. The display 232 displays an

image of the hard copy document itself and/or information necessary to process the information extracted off of the hard copy document.

The computer 230 is used to select portions of the stored document information contained in memory in accordance with content instructions which define portions of the stored document information required by an application unit. These content instructions may be provided by the application program. Alternatively, the content instructions can be inputted via an input device such as a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

The computer 230 is also used to format selected stored document information into the transmission format used by an application unit based on transmission format instructions. The transmission format instructions may be provided by the application program. Alternatively, the transmission format instructions can be inputted via a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

Thus, the computer 230 is used to generate an input file for a particular application unit. The computer 230 is connected to scanner memory 220, main, or permanent, memory 250, a printer 260, and application units 270, via a bus 240. Although FIG. 1 illustrates use of a bus to connect components together, it is understood that any routing or connecting link, implemented in hardware or software or both, can be employed instead of, or in addition to, a bus. Instructions to or in the computer 230 control the main memory 250, the printer 260, the application units 270, and the bus 240. Instructions to or in computer 230 can also control exchanges of information with scanner memory 220.

When the computer 230 generates an input file for a particular document, the computer 230 can send this input file directly to an application unit or can store this input file in the main memory 250 until required by an application unit. The main memory 250 may also optionally store a copy of the image information for the hard copy document and the textual information for the hard copy document. Thus, the image information and textual information from the hard copy document can be retrieved and printed out on printer 260. In addition, image and textual information stored in scanner memory 220 or in main memory 250 can be used to form additional input files at the time of input or at a later time, based on content instructions and transmission format instructions. Thus, the invention can, at the discretion of the user, eliminate the need to retain copies of hard copy documents, permitting a paperless office.

The application units 270 include particular application programs and devices which are controlled in accordance with information contained on hard copy document 100.

FIG. 2 illustrates an example of a hard copy document 100 which contains information to be processed by the instant invention. The document illustrated in FIG. 2 is a bill from XYZ Corporation to customer ABC Corporation. FIG. 2 is only an example of a type of document that can be processed by the instant invention.

In a first operational mode, the scanner 210 stores all of the information extracted off of hard copy document 100 in the scanner memory 220 or, alternatively, in main memory 250. The extracted information is stored in two forms. The actual image of the hard copy document 100 is stored as image information in the scanner memory 220. In addition, the scanner memory 220 stores textual information recognized on the hard copy document 100 by, for example, employing standard character recognition software. In the preferred embodiment, the textual information is stored in

ASCII format. The scanner memory 220 can be, for example, an electronic, magnetic, or optical memory.

FIG. 3A illustrates an enlarged view of the computer 230 of FIG. 1. This view will be used to describe a second mode of operation. In this second mode of operation, the hard copy document 100 is scanned and a copy of the document 100 is displayed on display 232 of computer 230, based on the contents of information temporarily stored in scanner memory 220. After the document is displayed on display 232, the computer 230 interactively prompts the user to identify the location of specific pieces of information on the hard copy document. In the FIG. 3A illustration, this prompt message is indicated as the message beginning with the arrow.

For example, the prompt message can ask the user to identify the location of account number information on the hard copy document. The user then uses an input device, such as keyboard 234 or mouse 236 or a touch screen, notepad, voice recognition device, or other input device to position a cursor on the display to identify the location of the information requested by the prompt message. For example, the cursor could be used to define a block (which could be highlighted) containing the requested information, followed by a mouse "enter" click. In this example, the user would move the mouse to identify the location of the account number information contained on the hard copy document 100. The computer 230 then stores the information which has been identified by the user as account number information in the appropriate address or subfile or as the appropriate variable or parameter or data field in memory. The computer then prompts the user to identify the location of other information on the hard copy document, such as, statement date information. The process proceeds until all of the desired information has been stored into the appropriate locations in memory.

FIG. 3B illustrates a variation of the second mode for interactively prompting the user for information. In FIG. 3B, the display is split into two portions. A left-hand portion 232L displays the image of the hard copy document and a right-hand portion 232R displays the required application program information. For example, in FIG. 3B, portion 232R displays a spreadsheet used by an application program. While observing the split display, the user can input instructions to associate specific pieces of information on the hard copy document (for example, the vendor name indicated by the mouse arrow 232A) with particular subfiles in memory (for example, the vendor field next to which the cursor 232C appears), using a mouse or other input device(s) or both. The split display also allows the user to generate content format instructions while observing the information required for a particular application program on the right-hand portion.

These second modes of operation are efficient for small businesses which receive a small number of a wide variety of invoices, since the user does not necessarily have to store all of the information that appears on the hard copy document. A further advantage is that data input is quicker, easier, and more accurate than with previous keyboard methodology. In addition, by specifying the location on the hard copy document of information, the user may optionally create a template, to be described in further detail below, for each different type of invoice. This template is stored for future use when another hard copy document in the same format is received.

More specifically, instructions from computer 230 can direct the scanner 210 and scanner memory 220, and/or main

memory **250**, to scan and/or store only specific portions of hard copy document **100**. After the interactive prompts required to obtain information for a desired application program, the unused information stored in scanner memory **220** or **250** can be erased. Further, scanning of a second identical document can be limited to only those portions of the document which contain needed information.

More specifically, in FIG. 2, the lines **10** drawn around certain portions of the document represent the areas which the user has previously identified as the portions of a document to be extracted by the scanner **210** and stored in scanner memory **220** and/or main memory **250**. Since the logo **20** and the message **30** have not been identified as an area to be scanned and stored, these areas are not scanned and stored in subsequent documents. Since the user has previously associated each of the areas **10** with a specific subfile of information, e.g., the account number, the scanned information is stored in memory locations corresponding to that subfile.

Data Processing

FIGS. **4-10** illustrate the flow of data in the FIG. 1 preferred embodiment. FIG. 4 illustrates the overall data flow for the FIG. 1 preferred embodiment. The preferred embodiment includes an input process module **1.0**, an information processing module **2.0**, and an output processing module **3.0**. The information processing module **2.0** is equipped to receive instructions from and transmit information to a user. The information processing module **2.0** can also transmit to and receive information from a remote external device through communication interface **4.0**. Input process module **1.0** and output processing module **3.0** can also access communication interface **4.0**. A module is implemented in hardware, software, or a combination of hardware and software. The specific implementation for a particular business application depends upon a variety of factors, for example, the relative costs of hardware and software implemented systems, the frequency with which a user will want to expand or modify the system, and the like.

FIG. 5 is a more detailed diagram of the input process module **1.0** of FIG. 4. The input process module **1.0** includes a character input module **1.1**, an image input module **1.2**, and, in the preferred embodiment, a character recognition device **1.3**. The character input module inputs textual information, such as alphanumeric characters, from an input device such as keyboard **234**. The image input module **1.2** inputs image information, for example, a digitized image of the actual appearance of hard copy document **100**. Textual information can, include textual input from an input device such as keyboard **234** and textual information extracted from the document by character recognition device **1.3**. Both types of information comprise an input document which is transmitted to information processing module **2.0**. In the FIG. 1 preferred embodiment, the processing performed by input process module **1.0** occurs in scanner memory **220**, computer **230**, and main memory **250**.

FIG. 6 illustrates information processing data flow for the FIG. 1 preferred embodiment, that is, FIG. 6 illustrates data flow in the information processing module **2.0**.

The information processing module **2.0** includes a maintain library module **2.1**, to be described in further detail below in conjunction with FIG. 7, a maintain definitions module **2.2**, to be described in further detail below in conjunction with FIG. 8, and a process document module **2.3** to be described in further detail below in conjunction with FIG. 9.

The information processing module **2.0** is the module which coordinates and drives the entire system. In the preferred embodiment, the information processing module **2.0** is implemented primarily by computer **230**.

FIG. 7 illustrates information processing data flow in the maintain library module **2.1**. The maintain library module **2.1** maintains a library of image information, for example, a digitized image representing the actual appearance of the hard copy document, and textual information of the hard copy documents for reference during processing. This library can be incorporated within scanner memory **220**, main memory **250**, or another independent memory, for example, a RAM disk. The maintain library module **2.1** includes a store document module **2.1.1**, a correct errors module **2.1.2**, a retrieve document module **2.1.3**, and a document file **2.1.4**. These modules operate collectively to store, retrieve, and correct document information.

The store document module **2.1.1**, prior to routing the document to the document file **2.1.4**, may provide information on recognition errors which may have occurred while inputting the document. For example, the store document module **2.1.1** identifies that a character contained on hard copy document **100** was not recognized. The store document module **2.1.1** also optionally causes a copy of the document and its parsing to be displayed on the display **232** for confirmation by the user. The user may utilize this opportunity to identify any errors in the displayed document and, in conjunction with the correct errors module **2.1.2**, to revise the document's parsing, if necessary, prior to storage of the document in memory. The module **2.1.1** also provides a facility for the user to name a particular hard copy document for cataloging, storage, and retrieval purposes. After the document is named, the store document module **2.1.1** stores copies of the document in the document file **2.1.4**.

The correct errors module **2.1.2** processes instructions from the user to correct errors identified by the store document module **2.1.1** and errors that have been spotted by the user during the confirmation process.

The retrieve document module **2.1.3** permits the user to retrieve a copy of a document previously stored in the document file **2.1.4**. As described above, long-term storage is provided by main memory **250**, if necessary.

FIG. 8 illustrates a more detailed information processing data flow diagram for the maintain definitions module **2.2** of FIG. 6. The maintain definitions module **2.2** allows the user to define system and document parameters and maintains the definitions of these system and document parameters. The maintain definitions module **2.2** includes a define template module **2.2.1** which allows the user to specify the location of information on the document. This information provided by the user defines a template which is used to extract information off the document and to associate the extracted information with a particular variable or subfile. These templates are illustrated by boxes **10** in the FIG. 2 example of a hard copy document. The maintain definitions module **2.2** can also access templates previously defined by the user and stored in main memory **250**. Templates can also be provided as part of software packages developed by program developers.

The maintain definitions module **2.2** also includes a define relationships module **2.2.2**. The define relationships module **2.2.2** allows the user to define data relationships, or logical relationships, between pieces of information extracted from the hard copy document. These pieces of information are then used to generate an input file for a selected computer application unit. The user defines these relationships by

content instructions. Alternatively, content instructions to define relationships can be provided by application software. If the user provides these content instructions, the content instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. Examples of content instructions, data, and logical relationships will be described in further detail in conjunction with FIGS. 11 and 12A, 12B, and 12C.

The maintain definitions module 2.2 also includes a define format module 2.2.3. The define format module 2.2.3 allows the user to define transmission formats for an input file which is then transmitted to a selected computer application unit. Selection of the transmission format of the input file is accomplished by the user through use of transmission format instructions. Alternatively, the applications software itself can generate its own transmission format instructions. When the user must specify transmission format instructions, the transmission format instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. A further description of various transmission formats will be provided below in conjunction with FIGS. 12A, 12B, 12C, 13A, 13B, and 13C.

A select definitions module 2.2.4 is also included in the maintain definitions module 2.2. The select definitions module 2.2.4 allows the user to store and select a set of definitions to be used for processing the document. The definitions identify pieces of information on the document by, for example, absolute location, variable location, or relative location, or by proximity to key words and/or symbols. These definitions are described in further detail below by way of an illustrative example.

FIG. 9 illustrates a more detailed information processing data flow diagram for the process document module 2.3. The process document module 2.3 processes the document after the document has been stored in the system. The process document module 2.3 gathers the appropriate information which has been stored, and creates input file(s) 2.3.3 for the selected application unit. The process document module 2.3 then transmits the input file(s) via bus 240 and/or communication interface 4.0 to an application unit 270, an output device such as printer 260, or to main memory 250.

The process document module 2.3 includes an extract data module 2.3.1. This module extracts data off of the document in accordance with the user's instructions, for example, the user-defined template, or through the interactive mode.

The process document module 2.3 also includes a preapplication process module 2.3.2 which gathers and associates information extracted from the document in accordance with content instructions. This module prompts the user for any additional information required to satisfy the relationships defined by the content instructions. The preapplication process module 2.3.2 also places the selected information into the transmission format defined by the transmission format instructions.

The preapplication process module 2.3.2 also generates the input file 2.3.3 for the selected application in accordance with the appropriate instructions. The input file 2.3.3 is then transmitted to bus 240 and/or communication interface 4.0 for transmission to a particular application unit 270.

FIG. 10 illustrates a detailed output data flow diagram for output module 3.0. Output module 3.0 outputs a textual and/or image copy of the document. In the FIG. 1 preferred embodiment, output module 3.0 is implemented by printer 260, associated software, and associated interface circuitry.

Operation

Examples of operation of a preferred embodiment will now be described.

The user enters the system by providing instructions to the information processing module 2.0. The user then instructs the information processing module 2.0 to conduct maintain library processing, maintain definitions processing, or process document processing.

If the user selects maintain library processing, the user then provides instructions to maintain or modify the document library through the maintain library module 2.1. For example, the user can direct the inputting and storage of a hard copy document 100 or can retrieve and output a document. The user requests inputting of a document through the store document module 2.1.1. The system then prompts the user to specify a storage location for the inputted document. The document is then read-in by the input process module 1.0. A textual copy and/or an image copy are stored into the document file 2.1.4. Errors which have occurred during inputting are identified and corrected by the correct errors module 2.1.2 and the user. The corrections are reflected in the document information stored in document file 2.1.4.

The retrieve document module 2.1.3 is used to retrieve and output a document. The system prompts the user to specify the storage location of a document and the type of document copy, for example, a textual or an image copy, to be outputted. The document is then outputted by the output process module 3.0.

If the user initially selected maintain definitions processing, the user would instruct the system to maintain and/or modify parameter definitions through the maintain definitions module 2.2. For example, the user can define and maintain a document template for extracting selected portions of information off of the hard copy document. The user can use the template to extract selected portions of information off of the hard copy document when the document is originally inputted, or alternatively, the user can use the template to identify selected portions of information for extraction off of an image copy of the document. In creating the template, the user identifies pieces of information on the document to be extracted and assigns a variable name, or subfile, to each piece of data.

The location of data to be extracted can be defined in a number of ways other than by use of a template. For example, the user can designate the absolute location of information on the document with respect to a grid overlaid on the document, e.g., always on line 3, starting in column 1. The user can also identify information by specifying the relative location of information to be extracted, e.g., always two lines below the piece of data named "salutation", starting in column 3. The user can also specify the location of information to be extracted by variable location specification. For example, if the hard copy document is a letter, the module would conduct a key word search for the term "Dear Sir:". Wherever this term "Dear Sir:" is located, this piece of data would be associated with the variable specified by the user, for example, the variable "salutation." In addition, a defined set of conventional symbols can be used to signify certain recurring data items for the convenience of users of the instant invention. For example, a "@" symbol can be used to delineate the vendor name as follows: "@XYZ Corporation@". Other examples of the use of symbols to delineate information will be described with reference to FIG. 14.

The maintain definition module 2.2 is also used to maintain data relationships in accordance with content instruc-

tions and to maintain input file formats in accordance with transmission format instructions. Relationships are defined and maintained between pieces of data, specified by, for example, the names of variables, through the define relationships module 2.2.2. The names of pieces of data on the document are retrieved by, for example, the define template module 2.2.1, and are passed to the define relationships module 2.2.2. The user may then provide any additional pieces of data needed to generate an input file for a particular application program or unit, such as an input file line number. The user, the applications software, and/or instructions previously stored in memory then establishes the contents of the input file by defining relationships between pieces of data using content instructions. Specific examples of content instructions will be discussed below in conjunction with FIGS. 11, 12A, 12B, 12C, 13A, 13B, and 13C.

The user and/or the applications software defines and maintains the transmission format of the input file to be used by a particular application program or unit through the define format module 2.2.3 in accordance with transmission format instructions. This is accomplished by defining the parameters to be used by the preapplication process module 2.3.2 in generating an input file. Parameters which would typically be required to generate an input file would include the character type, e.g., text or pixel; delimiters used between pieces of data, e.g., a slash or a semicolon; end of line characters, e.g., a carriage return or a line feed; and end of file characters. Examples of transmission formats will be described in further detail below in conjunction with FIGS. 11, 12A, 12B, 12C, 13A, 13B, and 13C.

If the user initially selected process document processing, the interface will then proceed to process the document through use of the process document module 2.3. For example, the user can extract specific portions of data from an image copy of a document, can generate an input file for transmission to an application program, or can directly process information interactively with an application program.

If the user desires to extract specific portions of data from an image copy of a hard copy document which has already been stored in memory, the user uses the extract data module 2.3.1 to identify a document to be processed. The document is then retrieved by the retrieve document module 2.1.3 and passed to the extract data module 2.3.1. The user can also select parameter definitions through the select definitions module 2.2.4.

The selected document template or parameter definition is passed to the extract data module 2.3.1. The extract data module 2.3.1 extracts pieces of data from the image copy of the document, as defined by the document template definition or the parameter definitions or both. This document data is then passed to preapplication process module 2.3.2.

The interface generates input file(s) 2.3.3 by use of the preapplication process module 2.3.2. The selected data relationship definition, as defined by the content instructions, and the selected record format definitions, as defined by the transmission format instructions, are passed to the preapplication process module 2.3.2. The preapplication process module 2.3.2 assembles the input file in accordance with the content instructions. The preapplication process module 2.3.2 also prompts the user for any additional pieces of data which need to be provided by the user. The input file is converted to the desired transmission format in accordance with the transmission format instructions. This physically formatted data is then stored in the input file 2.3.3.

The user can also use an application program to process information by loading the particular application program

into the computer 230 rather than by sending the input file to a remote application unit 270.

An illustrative example of the processing described above will now be described.

5 The user inputs instructions via keyboard 234 or another input device which indicate that the user desires to input and store a document. The computer 230 then prompts the user for the name of the document. In this example, the user desires to input the document of FIG. 2 and therefore names 10 the document "XYZ Corp. Bill 12/01/86." The computer then prompts the user to feed the hard copy document 100 into the scanner 210. The image of the hard copy document is displayed on display 232. The computer then prompts the user to identify the account number on the document. By use 15 of the mouse 236 or other input device to position a cursor on the display, the user indicates the location of the account number. The account number is then read-in to a subfile named "Account Number." This process proceeds until all of the desired information has been read-in and stored.

20 In this particular example, no errors were encountered while inputting the document. The user then directs that the document be stored for future reference in a document file.

Some time later, the user desires to retrieve and output the document and to generate input files based on information 25 from the document. The computer 230 prompts the user for the name of the document and the type of output. The user responds with "XYZ Corp. Bill 12/01/86" for a printed textual copy. The document is then retrieved from the document file and passed to the printer 260 for printing.

30 In order to generate an input file for a specific application program, the user selects the option to define a document template for use when each month's XYZ Corporation bill arrives. Accordingly, the user instructs the system to display 35 a copy of an XYZ Corporation bill on the display 232. The user then identifies pieces of data by absolute locations. That is, the user assigns specific names to information located at specific portions of the document. In this example, the user would input the following information:

40 Vendor-text, line 1, one line, column 1, 80 characters;
Account number-numeric, line 6, one line, column 25, 9 characters;
Statement date-date, line 9, one line, column 25, 8 characters;
45 Payment date-date, line 11, one line, column 25, 8 characters;
Previous balance-currency, line 7, one line, column 75, 9 characters;
New charges-currency, line 8, one line, column 75, 9 characters;
50 Other debits-currency, line 10, one line, column 75, 9 characters;
Finance charges-currency, line 12, one line, column 75, 9 characters;
Payments-currency, line 13, one line, column 75, 9 characters;
Other credits-currency, line 14, one line, column 75, 9 characters;
60 New balance-currency, line 15, one line, column 75, 9 characters.

The user also identifies data with variable locations. In this particular example, a variable location is specified as follows:

65 Heading 2-line, value="Mail To:"
The identification of Heading 2 as line information means that the system will search for occurrences of the character

string "Mail To:" and assign the line number which contains this character string to Heading 2.

The user also identifies data by relative locations. In this example, the user identifies the following relative location:

Mail To-text, Heading 2+1, 3 lines, column 60, 25 characters per line.

The instructions above instruct the system to assign the textual information beginning on one line after Heading 2 and continuing for 3 lines, in column 60, to the Mail To subfile.

As an alternative to inputting the actual line, column, and character numbers, the user can identify desired portions of the document by blocking, or highlighting, the desired portions using the mouse or other input device. In this case, the computer converts the highlighted portions into corresponding line, column, and character numbers.

FIG. 11 lists data corresponding to the hard copy document of FIG. 2 and the associated variable or subfile names.

Next, the user desires to define data relationships in accordance with content instructions. Examples of the type of contents which can be specified by a user are illustrated in FIGS. 12A, 12B, and 12C.

In this particular example, three separate departments of ABC Corporation require information from the XYZ Corporation bill. The first department requires vendor, account number, statement date, payment date, previous balance, new charges, debits, finance charges, payments, and new balance information. The second and third departments require mail to information and previous balance information. Each of these departments have their own application program which utilizes this information.

The user employs content instructions to designate how pieces of information, which have been extracted off of hard copy document 100, are directed to particular departments, that is, particular application programs. FIG. 12A illustrates the contents of the information to be transmitted to the first department. FIG. 12B illustrates the information to be transmitted to the second department. FIG. 12C illustrates the information to be transmitted to the third department. The content instructions, therefore, parse the information shown in FIG. 11 to various application programs, as shown by FIGS. 12A, 12B, and 12C. Content instructions can also be used to identify additional pieces of data which are required for the input files of the particular application programs. In this particular example, the specific application programs from the three departments all require numeric record number information, numeric horizontal position information, numeric vertical position information, and date received information. The horizontal and vertical position information is used by the application program to specify the location of the received information on a spreadsheet application program, in this example. The user may know in advance the content format required by each application program, that is, in this example, the location and type of information specified on the spreadsheet. The user may also employ the split display mode described with reference to FIG. 3B to generate content format instructions.

Using the content instructions, the user establishes the following contents for the input file corresponding to FIG. 12A:

Record number, horizontal position, vertical position, vendor;

Record number, horizontal position, vertical position, account number;

Record number, horizontal position, vertical position, statement date;

Record number, horizontal position, vertical position, date received;

Record number, horizontal position, vertical position, payment date;

Record number, horizontal position, vertical position, previous balance;

Record number, horizontal position, vertical position, new charges;

Record number, horizontal position, vertical position, 10 finance charges;

Record number, horizontal position, vertical position, payments;

Record number, horizontal position, vertical position, new balance.

15 Next, transmission format instructions are employed to define the transmission format of the input file for a specific application program or unit. FIG. 13A illustrates the transmission input file corresponding to FIG. 12A. FIG. 13B illustrates the transmission input file corresponding to FIG. 12B. FIG. 13C illustrates the transmission input file corresponding to FIG. 12C. A comparison of FIGS. 12B and 12C reveals that FIGS. 12B and 12C have the same contents. However, the information illustrated in FIG. 12B is being sent to a different application program than the information in FIG. 12C. These application programs require different transmission input formats as illustrated in FIGS. 13B and 13C. More specifically, the application program that receives the input file illustrated in FIG. 13B uses the greater than sign as a delimiter whereas the application program which receives the transmission input file shown in FIG. 13C uses a back-slash as the delimiter.

20 After the contents and the transmission format for the input file have been defined, and any additional information has been inputted, the input file is assembled and transmitted to the particular application program.

25 FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention. The hard copy document illustrated in FIG. 14 is first scanned and information from the hard copy document is stored into a memory. The interface 200 then identifies 30 portions of the hard copy document corresponding to various variables by recognizing a defined set of symbols. In the FIG. 14 example, triangles delineate the mailing address, 35 circles delineate the statement date, and squares delineate the new charges. Information from these portions of the hard copy document is stored in the corresponding memory locations or subfiles for each variable. The same set of symbols can be used to identify the same information from one document to the next. Thus, even if the physical formats of documents are not fixed from one document to the next, a diversity of hard copy documents can be processed without manually inputting data by recognition of the defined symbols.

40 45 50 55 Examples of readily available application programs are Quicken and Lotus 1, 2, 3 both of which are widely utilized in the business community. Quicken, for example is an easy-to-utilize program for writing checks and preparing business records. Payee, amount and address information may readily be transmitted from scanner memory 220 and/or main memory 250 to the Quicken application program for check writing functions and ledger keeping purposes. Lotus is a well known spreadsheet program which may process data input into specified cells once this data is placed in conventional Lotus format.

60 65 Thus, the instant invention provides an integrated and comprehensive system for handling information from a hard

copy document, thus permitting a paperless office. In addition, the invention permits data, extracted off of a hard copy document, to be easily manipulated into various logical and transmission formats required by a particular application unit. The invention also provides a low cost system for inputting information from a wide variety of hard copy documents into a memory.

The foregoing description has been set forth merely to illustrate preferred embodiments of the invention and is not intended to be limiting. Modifications are possible without departing from the scope of the invention.

For example, letters, checks, forms, pictures, reports, music scores, film, and other types of hard copy documents can be processed by the invention for accounts payable/receivable accounting, inventory control, record keeping, budgeting, data base management, music transcription, forms processing, computerized art, survey and questionnaire processing, statistical data analysis, correspondence processing and other applications.

Other automated digitizing units can be used in addition to or as an alternative to use of the scanner 210 as an input unit. Any electrical, magnetic, or optical device which extracts information off of a hard copy document, thereby eliminating the need to manually input significant amounts of information from the hard copy document is suitable for use as an automated digitizing unit. In addition, information can be input by user responses and digital and analog signals generated from various devices, and from computer files from other computer systems. Suitable hardware for inputting data includes a keyboard, a light pen, a mouse, a touch screen, a laser scanner, a microphone, a tablet, a disk drive, a magnetic tape drive, and a modem.

The interface 200 can also output information in forms other than a hard copy of textual or image information. For example, the interface 200 can output system responses, computer files, and digital and analog signals for transmission to other computer systems or to control systems. Suitable hardware for outputting information includes a disk drive, a magnetic tape drive, a cathode ray tube, a plasma screen, a printer, a plotter, a film developer, an amplifier, and a modem.

Since modifications of the described embodiment incorporating the spirit and substance of the invention may occur to persons skilled in the art, the scope of the invention should be limited solely with respect to the appended claims and equivalents.

What is claimed is:

1. A method of processing data on a computer comprising the steps of:

parsing a portion of data from a hard copy document corresponding to at least one field of information required by an application program;

converting said data to a format compatible with said application program; and

transmitting said formatted data to said at least one field of information of said application program;

said method to operate at the user's option in the following modes:

(a) receiving instructions from a user for parsing said portion of data;

(b) receiving instructions from a user to create a template for parsing said portion of data;

(c) selecting a template corresponding to said hard copy document; and

(d) utilizing said template for automatically parsing said portion of said data.

2. A method as recited in claim 1, wherein said converting step includes the step of recognizing alpha/numeric characters in said portion of said data using character recognition software.

3. A method as recited in claim 1, wherein said step of receiving instructions includes the step of displaying an image of at least said portion of data of said hard copy document on a portion of a display device and information corresponding to said field of information of said application program on another portion of a display device.

4. A method as recited in claim 1, wherein speech recognition techniques are used.

5. A method as recited in claim 1, wherein said user provides said instructions using a mouse control function.

6. A method as recited in claim 3, wherein said user supplies content instructions interactively to said computer, said content instructions defining said portion of data corresponding to at least one field of information required by an application program.

7. A method as recited in claim 1, wherein said converting step includes the step of recognizing image data in said portion of said data.

8. A method as recited in claim 1, wherein said user supplies content instructions interactively to said computer, said content instructions defining said portion of data corresponding to at least one field of information required by an application program.

9. An application program interface comprising:

a parser operable to select a portion of data from a hard copy document corresponding to at least one field of information required by an application program;

a converter operable to convert said data to a format compatible with said application program; and

a transmitter transmitting said formatted data to said at least one field of information of said application program;

said interface, at least partially implemented by a data processing device, and operable at the user's option in the following modes:

(a) receiving instructions from a user for parsing said portion of data;

(b) receiving instructions from a user to create a template for parsing said portion of data;

(c) selecting a template corresponding to said hard copy document; and

(d) utilizing said template for automatically parsing said portion of said data.

10. An application program interface as recited in claim 9, wherein said application program interface is adapted to operate in a distributed computing environment.

11. A multimode information processing system for inputting information from a hard copy document into at least one application program, comprising:

a parser operable to select a portion of data from said hard copy document corresponding to at least one field of information required by said application program;

a converter operable to convert said data to a format compatible with said application program; and

a transmitter transmitting said formatted data to said at least one field of information of said application program;

an interface, at least partially implemented by a data processing device, and operable at the user's option in the following modes:

(a) receiving instructions from a user for selecting said portion of data;

17

- (b) receiving instructions from a user to create a template for selecting said portion of data;
- (c) selecting a template corresponding to said hard copy document; and
- (d) utilizing said template for automatically parsing said portion of said data.

12. A system as recited in claim 11, wherein said user supplies content instructions interactively to said computer, said content instructions defining said portion of data corresponding to at least one field of information required by an application program.

13. A system as recited in claim 11, further comprising a speech recognition unit.

14. A system as recited in claim 12, wherein said system includes a speech recognition unit and, wherein said user provides said content instructions using at least one voice command.

15. A system as recited in claim 11, wherein said interface includes a display device which displays an image of at least said portion of data of said hard copy document on a portion of a display device and information corresponding to said field of information of said application program on another portion of a display device.

16. A multimode information processing system for inputting information from a hard copy document into at least one field of information required by an application program, comprising:

- an interface, including an electronic data processing device, and operable to parse a portion of data from said hard copy document corresponding to said at least one field of information required by said application program;
- a converter, including said electronic data processing device, and operable to convert said data to a format compatible with said application program; and
- a transmitter transmitting said formatted data to said at least one field of information of said application program;

said interface operable at the user's option in the following modes:

- (a) receiving instructions from a user for parsing said portion of data;
- (b) receiving instructions from a user to create a template for parsing said portion of data;
- (c) selecting a template corresponding to said hard copy document; and
- (d) utilizing said template for automatically parsing said portion of said data.

18

17. A multimode information processing system for inputting information from a hard copy document into at least one field of information required by an application program, comprising:

an interface, including at least in part an electronic data processing device, said interface operable in the following modes:

- (a) receiving instructions from a user for parsing a portion of data from said hard copy document corresponding to said at least one field of information;
- (b) receiving instructions from a user to create a template for parsing said portion of data;
- (c) selecting a template corresponding to said hard copy document; and
- (d) utilizing said template for automatically parsing said portion of said data; and

a transmitter transmitting said formatted data to said at least one field of information of said application program.

18. A multimode information processing system as recited in claims 11, 16 or 17, wherein said system is a distributed system.

19. A method for multimode information data processing for inputting information from a hard copy document into at least one field of information required by an application program, comprising the steps of:

receiving input data corresponding to at least a portion of said hard copy document into a data processing device, said data processing device operable in the following modes:

- (a) receiving instructions from a user for parsing a portion of said input data from said hard copy document corresponding to said at least one field of information required by said application program;
- (b) receiving instructions from a user to create a template for parsing said portion of data;
- (c) selecting a template corresponding to said hard copy document; and
- (d) utilizing said template for automatically parsing said portion of said data; and

transmitting said formatted data to said at least one field of information of said application program.

20. A method as recited in claims 1 or 19, wherein said data processing method comprises a distributed computing method.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTIONPATENT NO. : 6,094,505
DATED : July 25, 2000
INVENTOR(S) : Robert Lech, Mitchell A. Medina and Catherine B. Elias

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 15,

Line 61, after "portion of data" insert -- without respect to a fixed location of said portion of data --.

Column 16,

Lines 42 and 67, after "portion of data" insert -- without respect to a fixed location of said portion of data --.

Column 17,

Line 43, after "portion of data" insert -- without respect to a fixed location of said portion of data --.

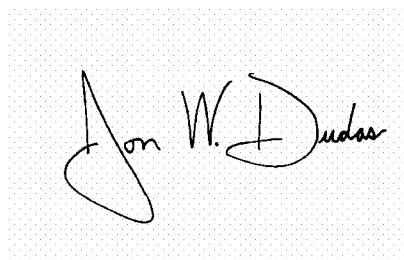
Column 18,

Line 11, after "field of information" insert -- without respect to a fixed location of said portion of data --.

Line 36, after "said application program" insert -- without respect to a fixed location of said portion of said input data --.

Signed and Sealed this

Twenty-first Day of June, 2005

A handwritten signature in black ink, appearing to read "Jon W. Dudas". The signature is written in a cursive style with a large, stylized "W" and "D".

JON W. DUDAS
Director of the United States Patent and Trademark Office

(12) **United States Patent**
Lech et al.(10) Patent No.: US 6,683,697 B1
(45) Date of Patent: Jan. 27, 2004(54) **INFORMATION PROCESSING
METHODOLOGY**(75) Inventors: **Robert Lech**, Jackson, NJ (US);
Mitchell A. Medina, Essex Fells, NJ
(US); **Catherine B. Elias**, Plainsboro,
NJ (US)(73) Assignee: **Millenium L.P.**, George Town (KY)(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 0 days.(21) Appl. No.: **09/458,162**(22) Filed: **Dec. 9, 1999****Related U.S. Application Data**(63) Continuation of application No. 09/044,159, filed on Mar.
19, 1998, now Pat. No. 6,094,505, which is a continuation
of application No. 08/487,150, filed on Jun. 7, 1995, now
Pat. No. 5,768,416, which is a division of application No.
08/348,224, filed on Nov. 28, 1994, now Pat. No. 5,625,465,
which is a continuation of application No. 08/143,135, filed
on Oct. 29, 1993, now Pat. No. 5,369,508, which is a
continuation of application No. 07/672,865, filed on Mar.
20, 1991, now Pat. No. 5,258,855.(51) **Int. Cl.⁷** **G06F 15/40**(52) **U.S. Cl.** **358/1.15; 382/180**(58) **Field of Search** **358/1.15; 382/175,
382/177, 180, 282, 287, 306, 317**(56) **References Cited****U.S. PATENT DOCUMENTS**

3,013,718 A	12/1961	Shepard et al.
3,200,372 A	8/1965	Hamburg
3,303,463 A	2/1967	Hamburg
3,434,110 A	3/1969	Bucklin, Jr. et al.
3,492,653 A	1/1970	Fosdick et al.
3,582,883 A	6/1971	Shepard
3,582,884 A	6/1971	Shepard

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

EP	0 107 083 B1	7/1988
JP	64-38883	2/1989
JP	3-161886	7/1991

OTHER PUBLICATIONS

Que's Computer User's Dictionary, 2nd Ed., Bryan Pfaffenberger (author); 1991; p. 144.

"Kurzweil 5200 Intelligent Scanning System", Xerox Imaging Systems, Inc., 1990.

PC Magazine, vol. 5, No. 16, Sep. 30, 1986.*TopScan Professional User's Guide*, Calera Recognition Systems, p. v-vii, xi-xii, and 1-6, 1989.Edward O. Welles, *Decisions, Decisions*, Inc. , Aug. 1990,
pp. 80-90.

(List continued on next page.)

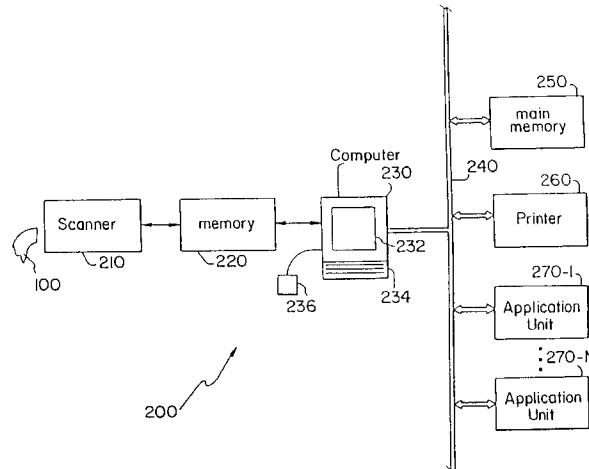
Primary Examiner—Thomas D. Lee

Assistant Examiner—Stephen Brinich

(74) Attorney, Agent, or Firm—Foley & Lardner

ABSTRACT

An information processing methodology gives rise to an application program interface which includes an automated digitizing unit, such as a scanner, which inputs information from a diversity of hard copy documents and stores information from the hard copy documents into a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which designate portions of the stored document information required by a particular application program. The selected stored document information is then placed into the transmission format required by a particular application program in accordance with transmission format instructions. After the information has been transmission formatted, the information is transmitted to the application program. In one operational mode, the interface interactively prompts the user to identify, on a display, portions of the hard copy documents containing information used in application programs or for storage.

101 Claims, 15 Drawing Sheets

U.S. PATENT DOCUMENTS

3,584,144 A 6/1971 Shepard et al.
 3,631,396 A 12/1971 Spertus
 3,832,682 A 8/1974 Brok et al.
 3,848,228 A 11/1974 MacNeill
 3,903,517 A 9/1975 Hafner
 RE29,104 E 1/1977 Shepard
 4,021,777 A 5/1977 Shepard
 4,034,343 A 7/1977 Wilmer 340/146.3 MA
 4,041,454 A 8/1977 Shepard et al.
 4,047,154 A 9/1977 Vitols et al.
 4,132,978 A 1/1979 Mercier
 4,387,964 A 6/1983 Arrazola et al.
 4,564,752 A 1/1986 Lepic et al.
 4,572,962 A 2/1986 Shepard
 4,593,367 A 6/1986 Slack et al.
 4,659,940 A 4/1987 Shepard
 4,667,248 A 5/1987 Kanno 358/280
 4,672,678 A 6/1987 Koetzuka et al.
 4,760,246 A 7/1988 Shepard
 4,760,606 A 7/1988 Lesnick et al. 382/48
 4,782,509 A 11/1988 Shepard
 4,802,104 A 1/1989 Ogiso 364/518
 4,802,231 A 1/1989 Davis
 4,803,734 A 2/1989 Onishi et al.
 4,931,957 A 6/1990 Takagi et al. 364/521
 4,933,979 A 6/1990 Suzuki et al. 381/61
 4,974,260 A 11/1990 Rudak
 5,017,763 A 5/1991 Shepard
 5,031,121 A 7/1991 Iwai et al. 364/523
 5,034,990 A 7/1991 Klees 382/22
 5,052,038 A 9/1991 Shepard
 5,095,445 A 3/1992 Sekiguchi 364/514
 5,140,139 A 8/1992 Shepard
 5,140,650 A 8/1992 Casey et al. 382/61
 5,153,927 A 10/1992 Yamanari 382/61
 5,159,667 A 10/1992 Borrey et al.
 5,191,525 A 3/1993 LeBrun et al.
 5,218,539 A * 6/1993 Elphick et al. 707/513
 5,228,100 A 7/1993 Takeda et al. 382/61
 5,245,166 A 9/1993 Shepard
 5,251,268 A 10/1993 Colley et al.
 5,257,328 A 10/1993 Shimizu
 5,258,855 A 11/1993 Lech et al. 358/462
 5,282,267 A 1/1994 Woo, Jr. et al.
 5,367,619 A 11/1994 Dipaolo 395/149
 5,404,294 A 4/1995 Karnik 364/419.1
 5,416,849 A 5/1995 Huang 382/173
 5,444,840 A 8/1995 Froessl
 5,448,738 A 9/1995 Good et al. 395/700
 5,452,379 A 9/1995 Poor 382/317
 5,455,875 A 10/1995 Chevion et al.
 5,506,697 A 4/1996 Li et al. 358/448
 5,511,135 A 4/1996 Rhyne et al.
 5,526,447 A 6/1996 Shepard
 5,550,930 A 8/1996 Berman et al.
 5,555,325 A 9/1996 Burger
 5,696,854 A 12/1997 Shepard
 5,734,761 A 3/1998 Bagley
 5,852,685 A 12/1998 Shepard
 5,923,792 A 7/1999 Shyu et al.
 5,933,531 A 8/1999 Lorie
 6,094,505 A 7/2000 Lech et al.

OTHER PUBLICATIONS

Palantir PagePro User's Guide, Rev. A, Dec. 1986, sections 1 and 4 and Appendix C.
 PageRead Library Developer's Guide, Rev. B, Aug. 1989, pp. PRL1-3, SHC25-26, TUT9-10.

With AutoClass there is no more to Indexing than OCR, Remittance and Document Processing Today.
 Daniel Borrey, *Machine Recognition and Classification of Documents*, Remittance and Document Processing.
The Very Best in Optical Character Recognition, Imaging, Mar. 1992, pp. 43-47.
Why Insurance Companies Take the Risk on Document Imaging, Imaging, Mar. 1992, pp. 48-54.
 "OCR for Forms" (advertisement), Imaging, Apr. 1992.
 David Black, *The Right and The Wrong Ways to Index, Imaging*, May 1992, pp. 47-50.
 Greg Bartels, *How to Successfully Convert Your Backfiles, Imaging*, May 1992, pp. 55-56.
But, Is It a Boy or Girl ? Imaging, Oct. 1992, p. 10.
 Gerry Frieser, *Suddenly, OCR is a "Must Buy"*, Imaging, Dec. 1992, pp. 22-25.
 Gerry Frieser, *Suddenly, OCR is a "Must Buy"*, Imaging, Dec. 1992, pp. 22-25.
How Form Processing Works, Plus Pros and Cons, Imaging, Dec. 1992, p. 36.
Forms Processing Products Meet the Challenge of OCRing Forms, Imaging, Dec. 1992, p. 38-40.
 Herbert F. Schantz, *Forms Automation and Integrated Imaging (OCR) Systems*, Remittance and Document Processing Today, Mar.-Apr. 1991, pp. 9-11.
 Don Merz, *OCR: A Health Insurance Applications, Remittance and Document Processing Today*, Jul.-Aug. 1989, p. 18-20.
Industry News, Remittance and Document Processing Today, Jul-Aug. 1989, p. 22.
Industry News and New Products, Remittance and Document Processing Today, Oct. 1984, pp. 17-18.
 R. C. Gonzalez, *Designing Balance into An OCR System*, Remittance and Document Processing Today, Mar. 1988, pp. 7, 10-11.
 Ambrose R. Rightler, *OCR Quality Control Procedures for Remittance Processing: Can You Afford to be Without It?* Remittance and Document Processing Today, Mar. 1988, pp. 12-15.
Product Watch, MacWeek, Oct. 3, 1989, pp. 32, 34, 38, 40, 42.
Industry News, Remittance and Document Processing Today, Jan. 1989, p. 9.
 Scott Beamer, *Mac OCR Takes a Big Step Forward*, MacWeek, Jun. 13, 1989.
 Matthew Lake, *Strength of Character (Recognition)*, Publish, Jan. 1991, pp. 62-67.
 R. David Nelson and Karen A. Hamill, *Optical Scanning at Chemical Abstracts Service for Building Computer Files From Printed Index Data*, Recognition Technologies Today, Feb. 1985, pp. 1-6, 15.
 Gerald Farmer, *HNC IDEPT™ and Recognition Enhanced Data Entry: The Cost-Cutting Approach to Automated Data Entry*, Remittance and Document Processing Today, Jan.-Feb. 1991, pp. 24-26.
 David Gertler, *Automated Data Entry*, Seybold Report on Desktop Publishing, Jan. 15, 1990, pp. 3-17.
 Eric Aas and Peter Davidoff, *Teaching Your Scanner to Read*, Personal Publishing, May 1990, pp. 28, 31, 33.
 Phillip Robinson et al., *Character Witness*, MacUser, Jul. 1990, pp. 120-136.
 Brita Meng, *Text Without Typing*, MacWorld, Oct. 1990, pp. 177-183.

Jim Heid, *Getting Started with Optical Character Recognition*, MacWorld, Oct. 1990, pp. 297–301.
Stanford Diehl and Howard Eglowstein, *Tame the Paper Tiger*, Byte, Apr. 1991, pp. 220–238.
Alan Joch and Rich Graham, *Voices of Experience*, Byte, Apr. 1991, pp. 239–241.
Gregory Boleslavsky and Roman Tutunikov, *The New Generation of OCR*, Inform, Jan. 1990, pp. 34–37.
Calera Recognition Systems, Inc., TopScan Professional User's Guide: Complete Document Recognition for PCs and Compatibles (1989).
Calera Recognition Systems, Inc., TopScan Professional Installation Notes for Scanners, Fax Cards, and System Configuration (1989).
Calera Recognition Systems, Inc., TopScan Professional Troubleshooting Guide (1998).
Xerox Imaging Systems, Inc., Datacopy AccuText User's Guide (1989).

Invoice from Corporate Intelligence Corporation to Workman, Nydegger & Seeley, dated Sep. 21, 1999.

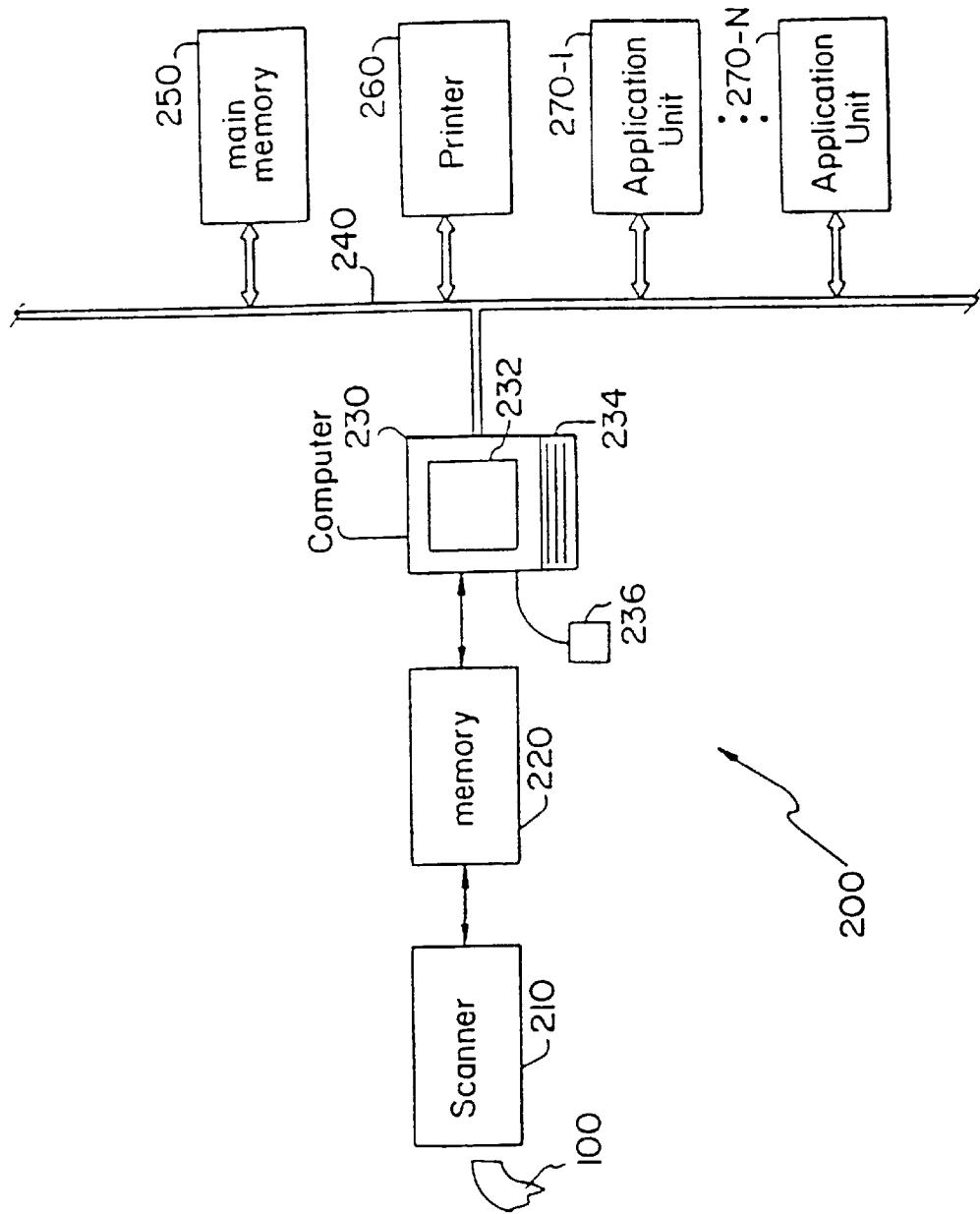
Examiner Interview Summary Record, Jan. 25, 1996, and related Amendment, for Application Ser. No. 08/097,131.

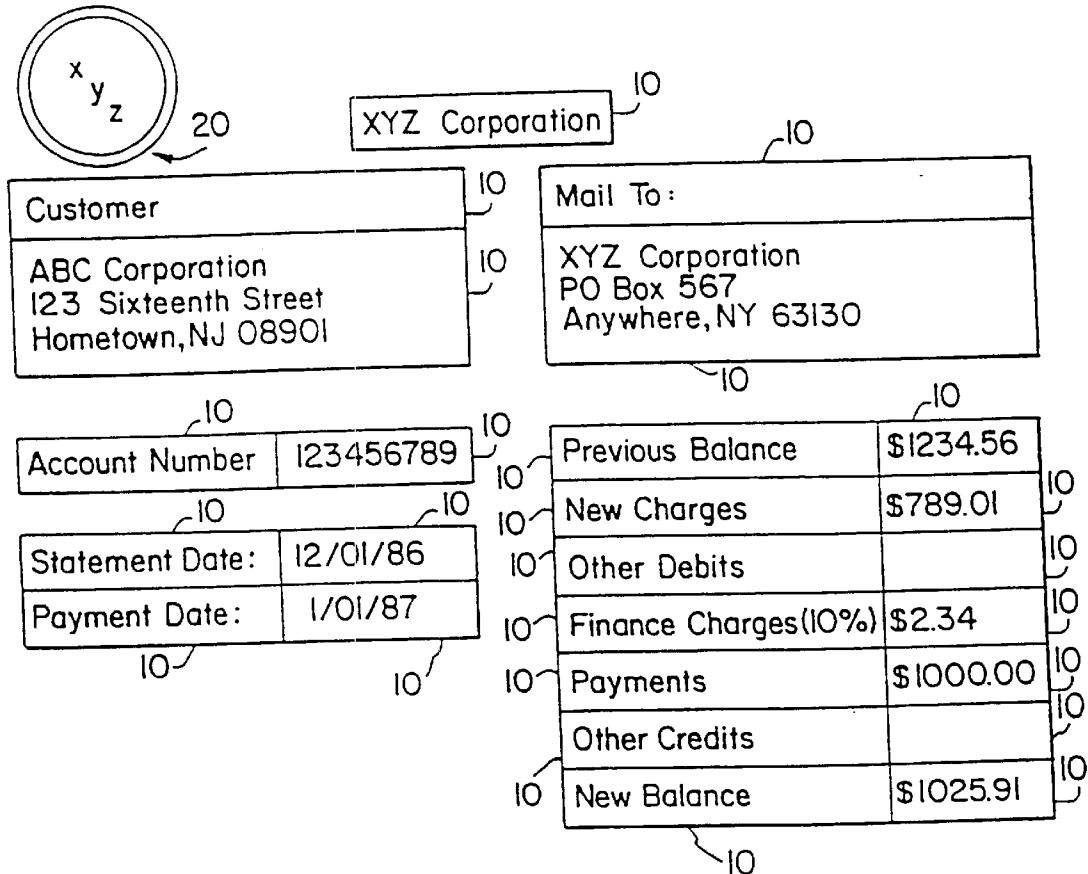
Jane B. Newman, Formstar Ad; “Stack the Facts, Not the Forms” –May 24, 1987; “Stack and Send Just the Facts –and Improve the Efficiency of your Forms Application”.

TeleImage Systems Document and Image Database Systems User's Manual –Ramat Gan, Israel; Table of Contents and pp. 2–1 through 5–20.

Form Out! Programmer's Manual; TeleImage Systems, Ramat Gan Israel; published Feb. 1991 pp. I, II, III, IV, V, VI, VII; 1.1–6.29 and A.1–G.11.

* cited by examiner





Winter Sale In Effect Throughout January!

30

FIG. 2

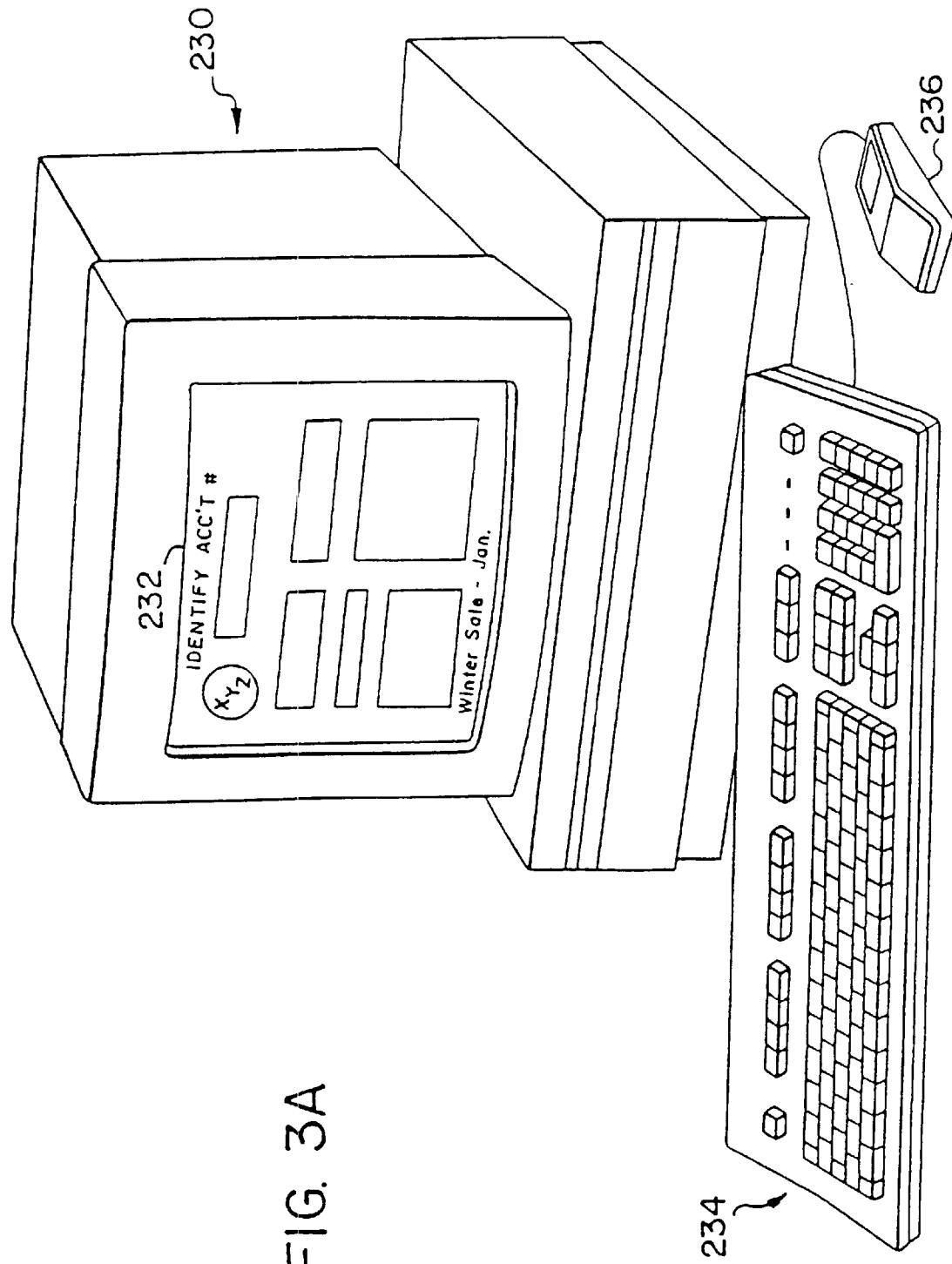


FIG. 3A

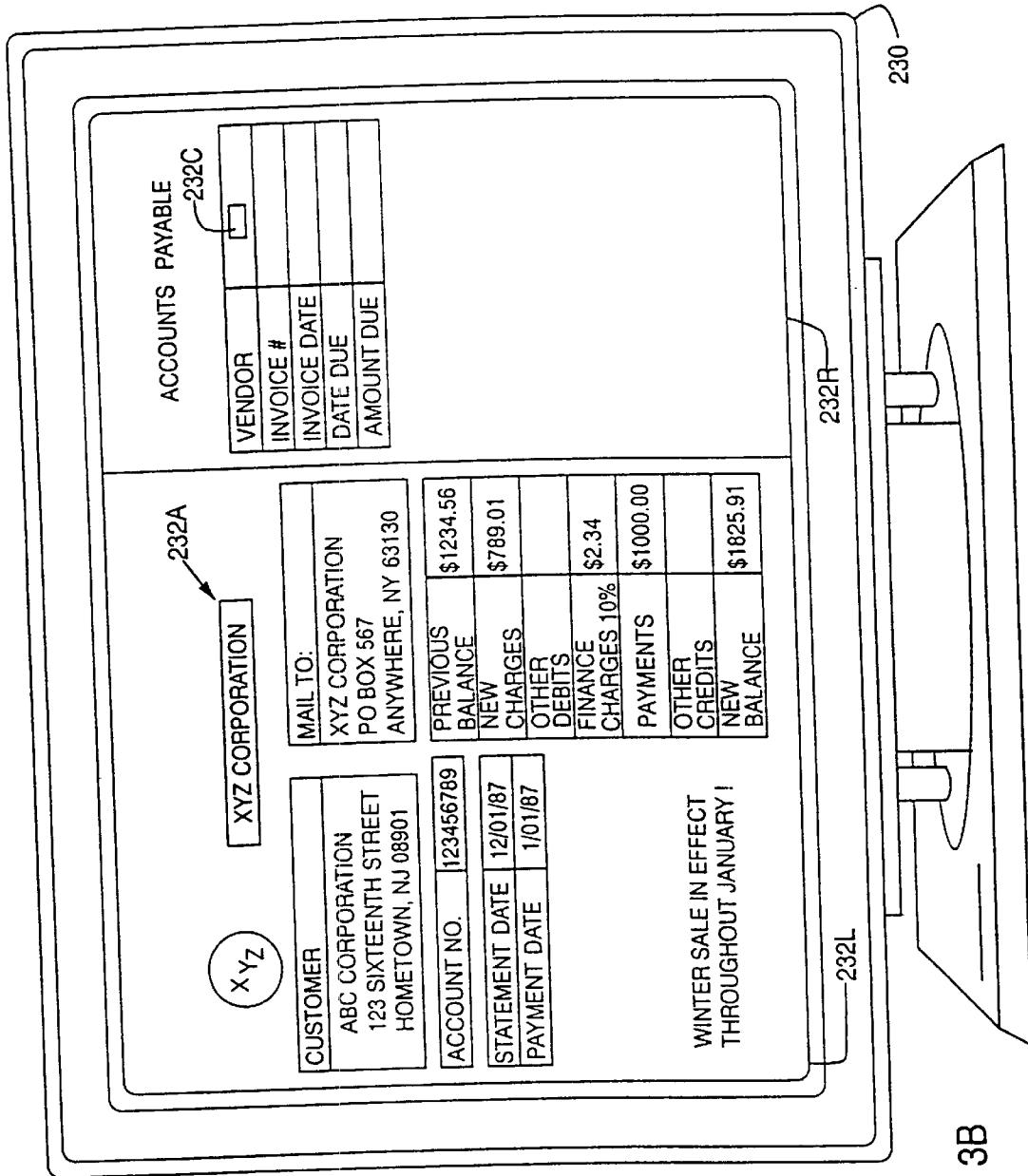


FIG. 3B

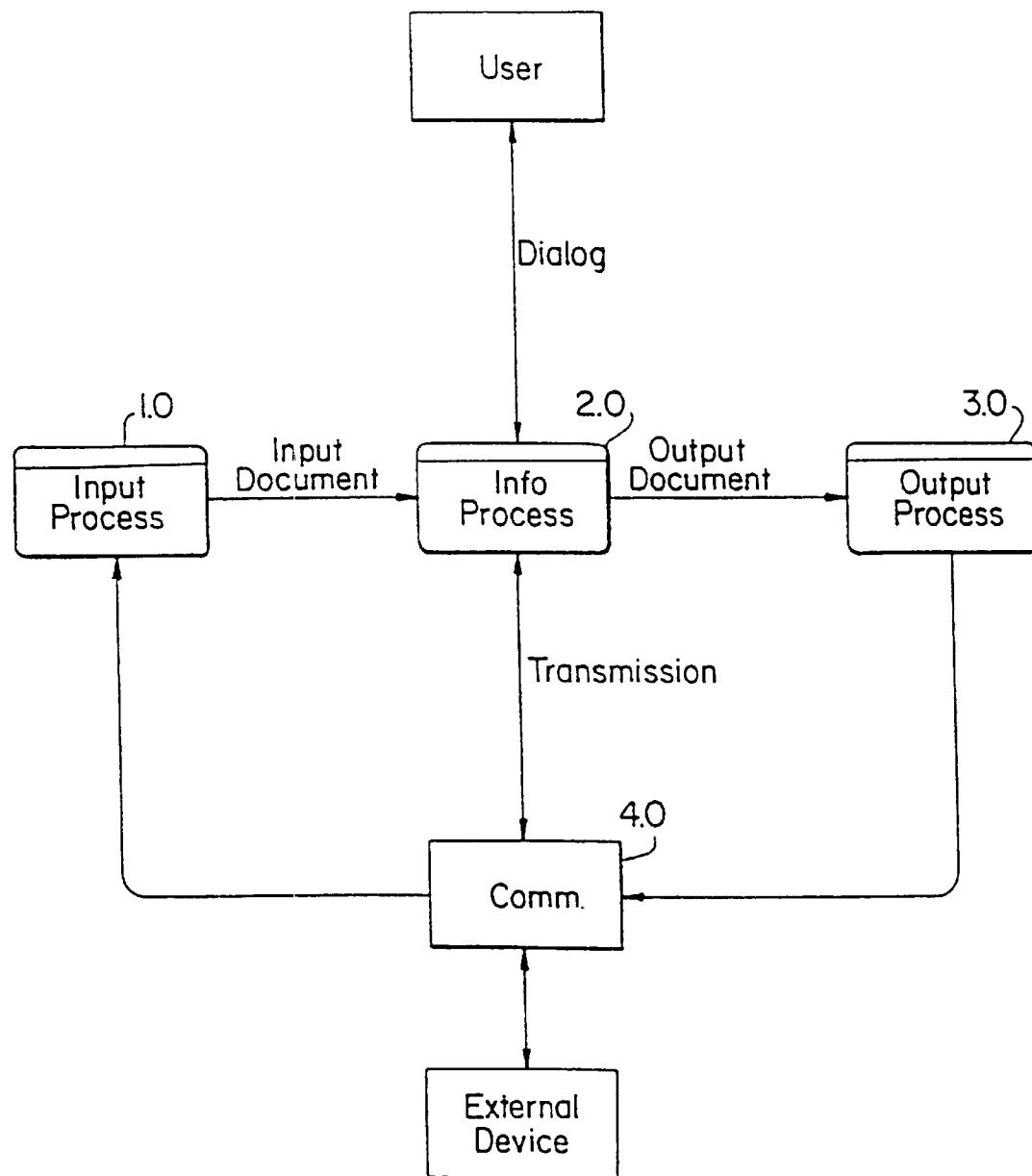
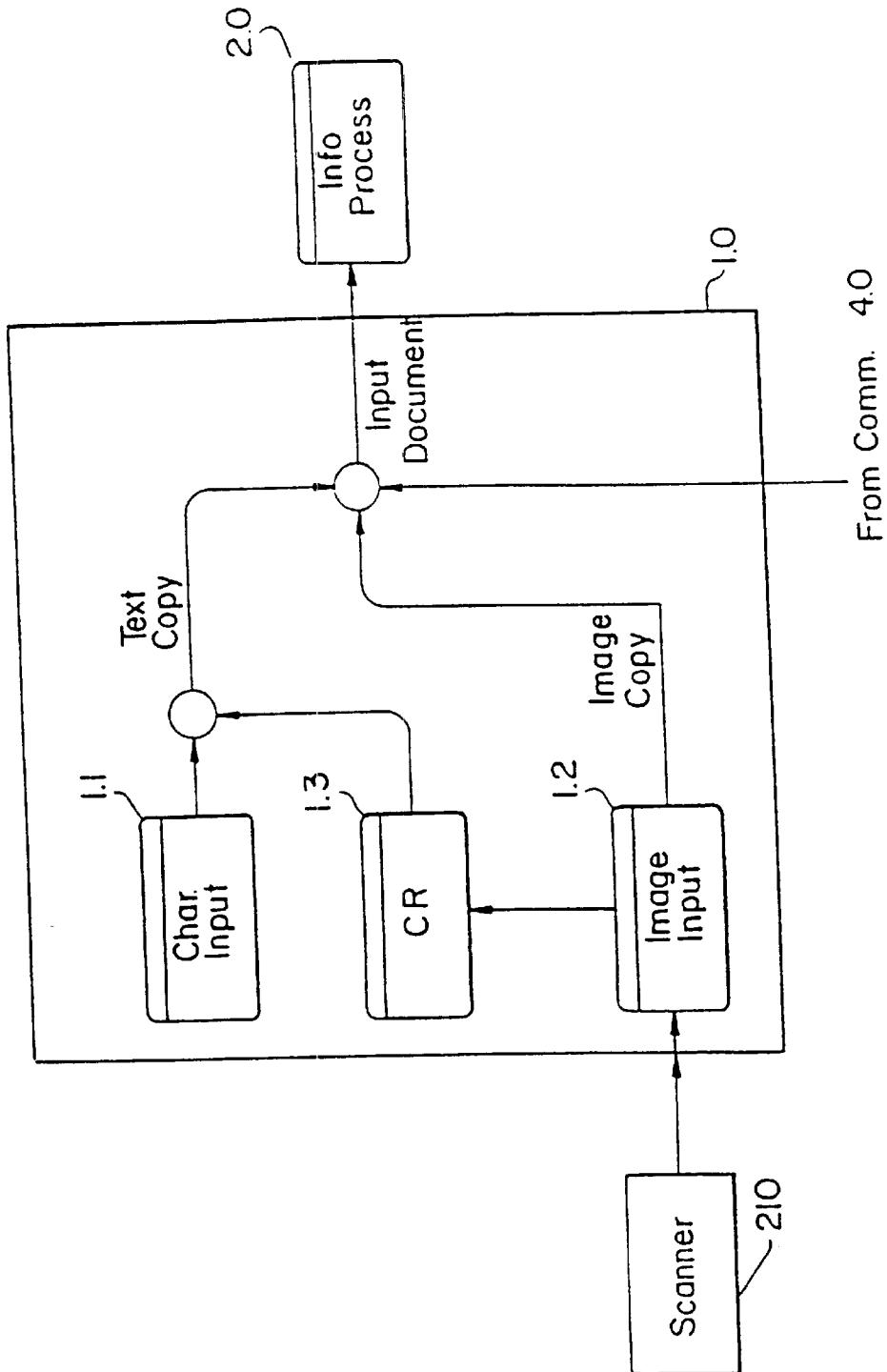


FIG. 4



५६

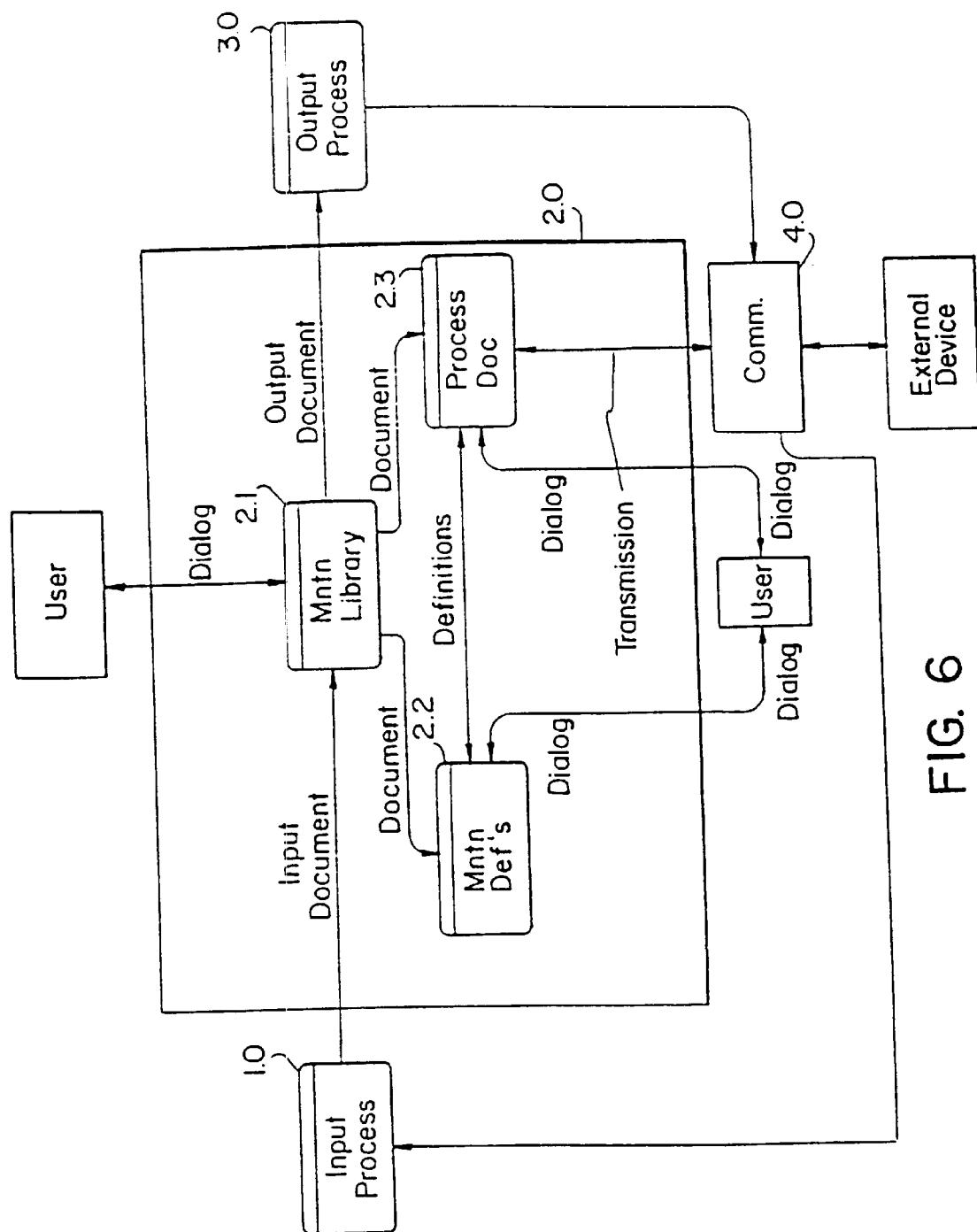


FIG. 6

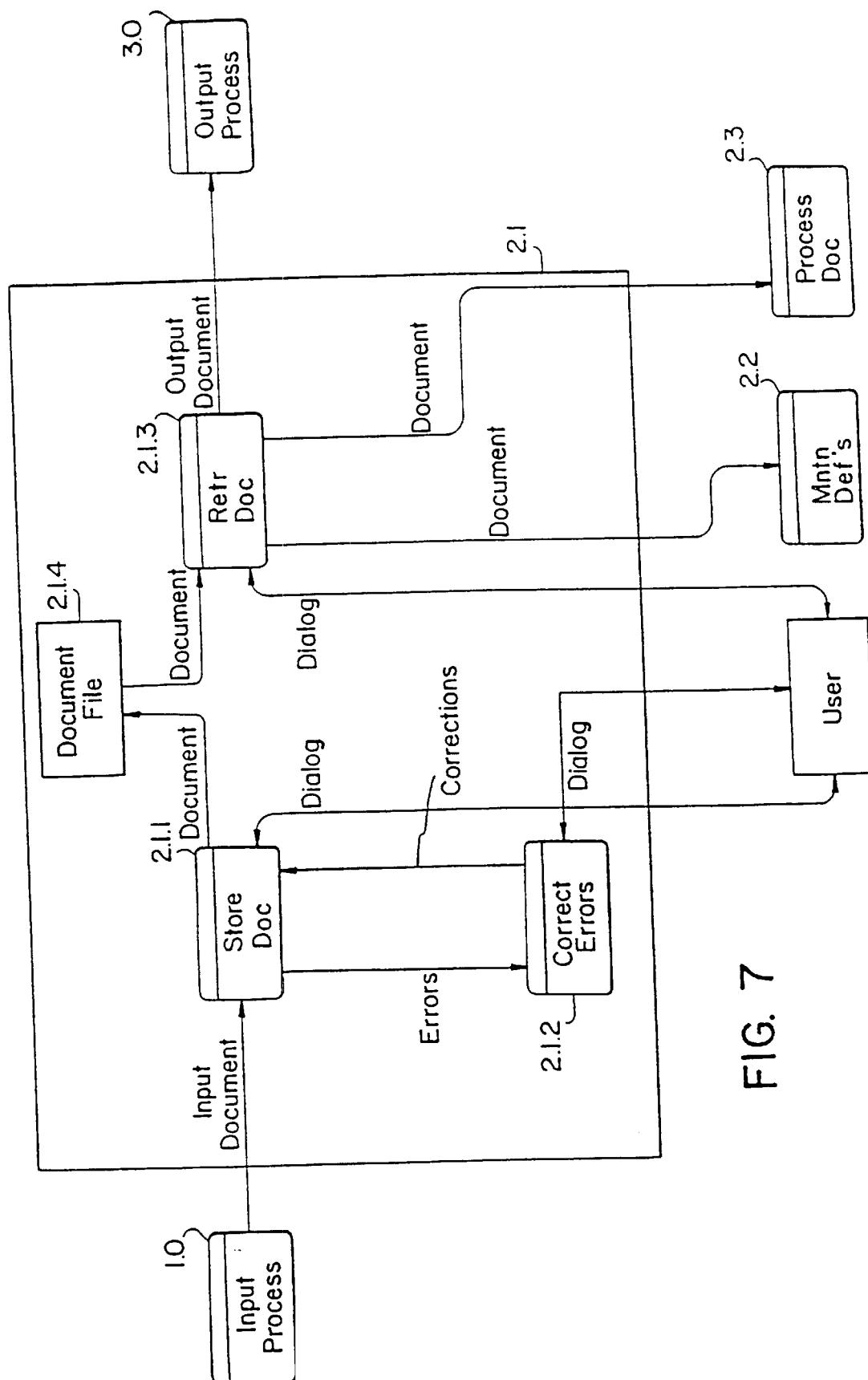


FIG. 7

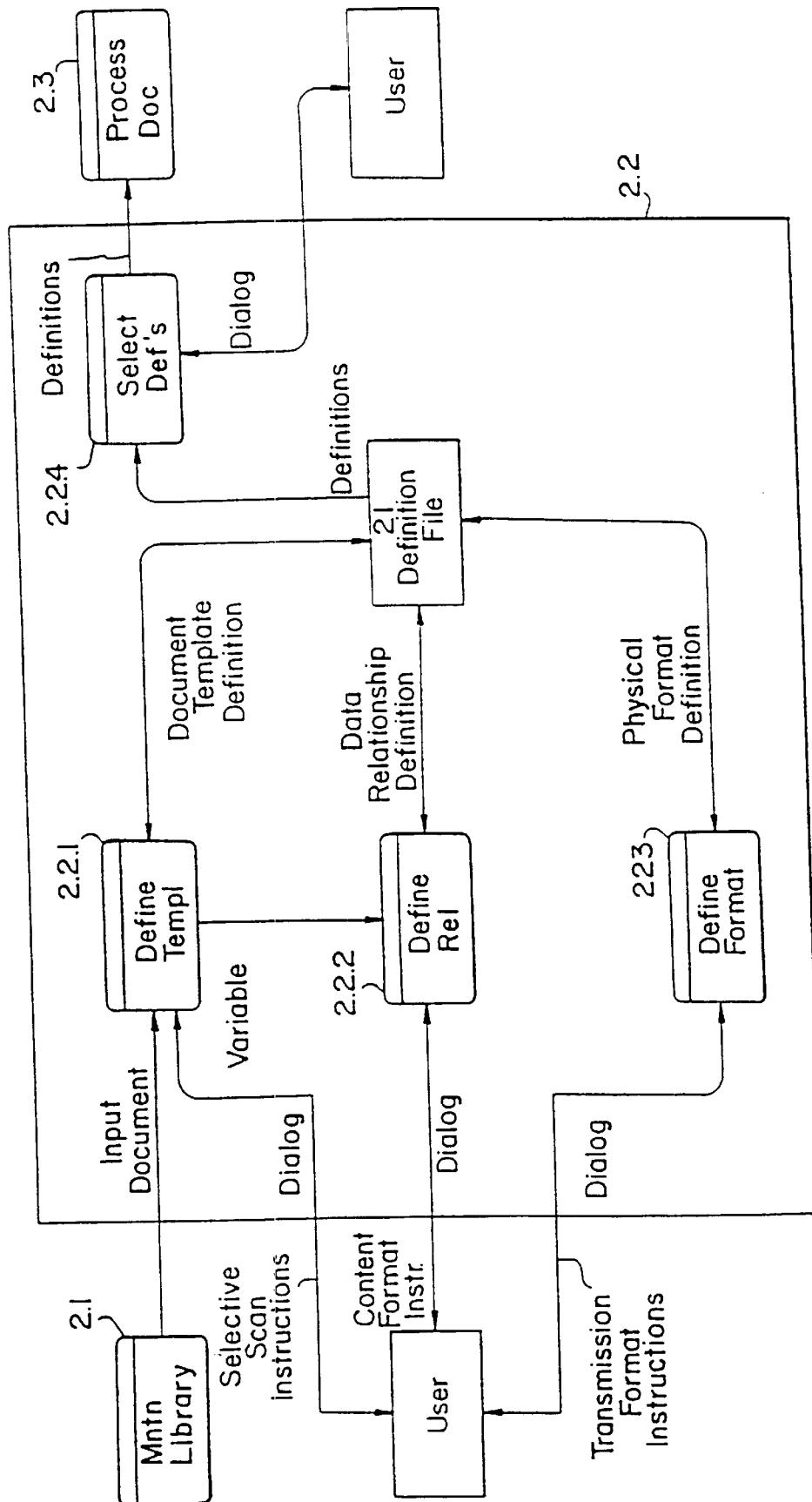


FIG. 8

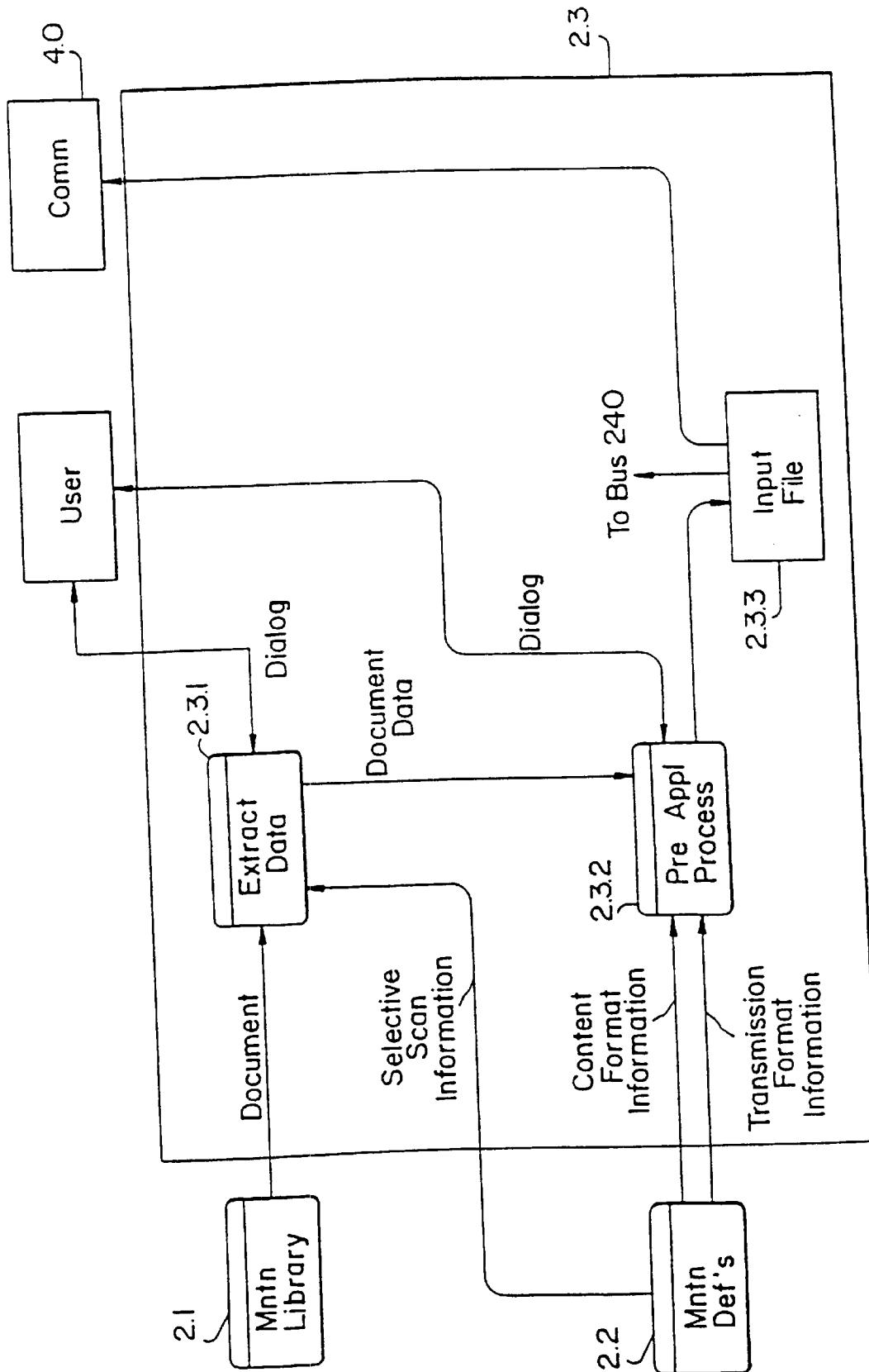


FIG. 9

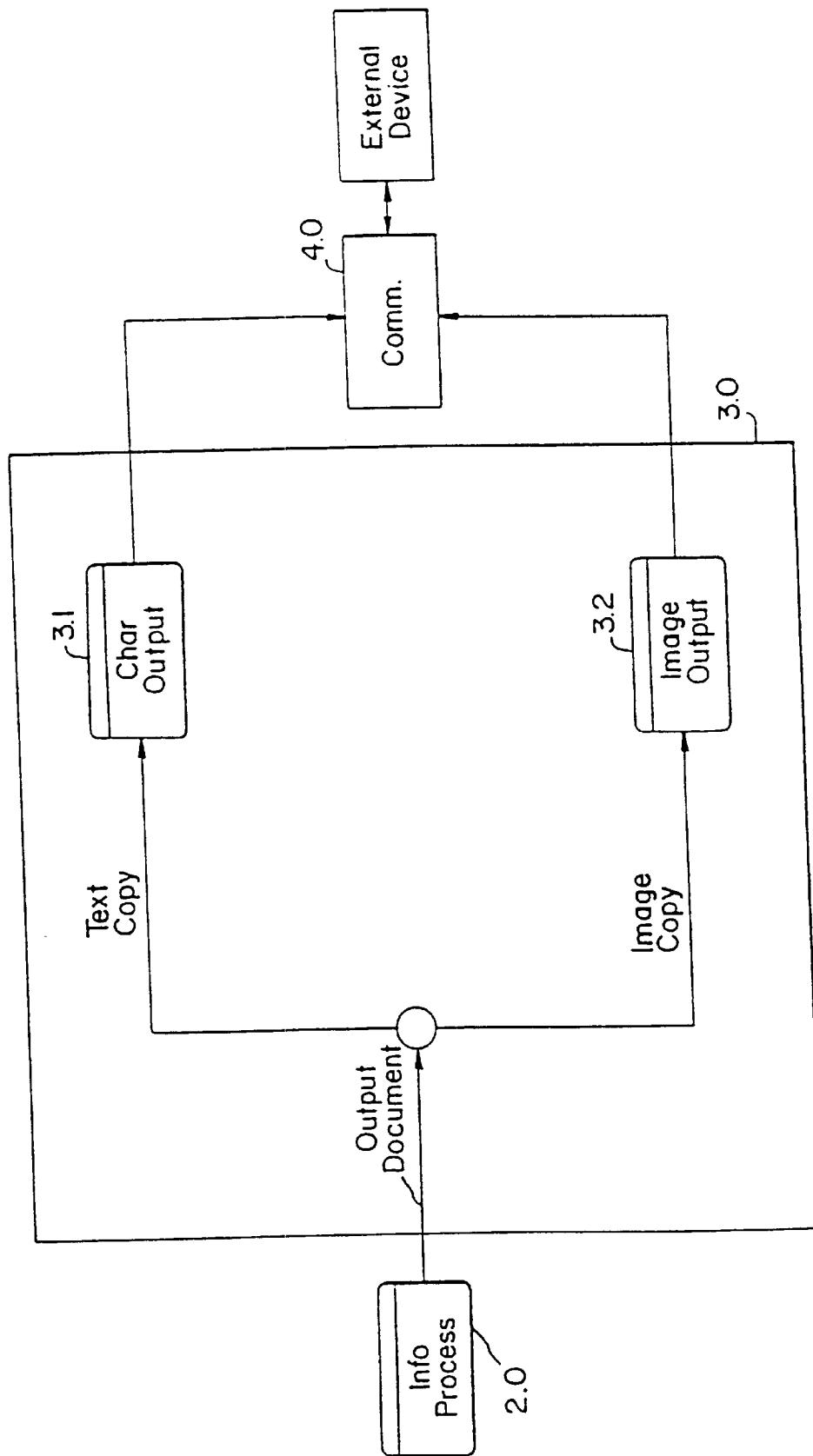


FIG. 10

FIG. 11

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Heading 2	2
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
Other Credits	
New Balance	\$1025.91

FIG. 12A

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
New Balance	\$1025.91

FIG. 12B

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

FIG. 12C

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

FIG.13A

{ > 1 > 1 > 1 > "XYZ Corporation"
> 2 > 2 > 25 >+ 123456789 >
> 3 > 2 > 1 > D12 / 01 / 86 >
> 4 > 2 > 11 > D12 / 15 / 86 >
> 5 > 2 > 21 > D01 / 01 / 87 >
> 6 > 10 > 25 > \$1234.56 >
> 7 > 11 > 25 > \$789.01 >
> 8 > 13 > 25 > \$2.34 >
> 9 > 14 > 25 > \$1000.00 >
> 10 > 16 > 25 > \$1025.91 >

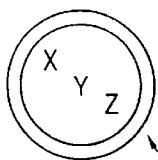
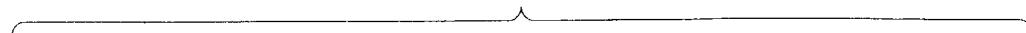
FIG.13B

{ > 1 > 1 > 1 > \$1234.56 >
> 2 > 2 > 1 > "XYZ Corporation"
> 3 > 3 > 1 > "PO Box 567"
> 4 > 4 > 1 > "Anywhere, NY 63130"

FIG.13C

{ / 1 / 1 / \$1234.56 //
/ 2 / 2 / 1 / *XYZ Corporation*
/ 3 / 3 / 1 / *PO Box 567*
/ 4 / 4 / 1 / *Anywhere, NY 63130*

FIG. 14



XYZ Corporation

Customer

ABC Corporation
123 Sixteenth Street
Hometown, NJ 88981

Mail To:

△ XYZ Corporation
PO Box 567
Anywhere, NY 63130 △

Account Number	123456789	Previous Balance	\$1234.56
Statement Date:	○ 12/01/86 ○	New Charges	<input type="checkbox"/> \$789.01 <input type="checkbox"/>
Payment Date:	1/01/87	Other Debits	
		Finance Charges (10%)	\$2.34
		Payments	\$1000.00
WINTER SALE IN EFFECT THROUGHOUT JANUARY:		Other Credits	
30		New Balance	\$1025.91

1

INFORMATION PROCESSING
METHODOLOGY

This application is a continuation of application Ser. No. 09/044,159, filed Mar. 19, 1998 (now U.S. Pat. No. 6,094,505), which is a continuation of application Ser. No. 08/487,150, filed Jun. 7, 1995 (now U.S. Pat. No. 5,768,416), which is a divisional of Ser. No. 08/348,224, filed Nov. 28, 1994 (now U.S. Pat. No. 5,625,465), which is a continuation of Ser. No. 08/143,135, filed Oct. 29, 1993 (now U.S. Pat. No. 5,369,508), which is a continuation of Ser. No. 07/672,865, filed Mar. 20, 1991 (now U.S. Pat. No. 5,258,855).

BACKGROUND OF THE INVENTION

The invention is directed to a system for efficiently processing information originating from hard copy documents. More specifically, the invention is directed to a hard copy document to application program interface which minimizes the need to manually process hard copy documents.

In the past, information contained on hard copy documents was manually entered into a computer via the input controller of a particular computer. The original document was then filed away for future reference. Automatic input of data was limited to the input of Magnetic Ink Character Recognition (MICR) data and to Optical Character Recognition (OCR) data. This fixed-position data was forwarded directly to a dedicated computer application specifically designed to accommodate the input format. In more recent years, typewritten text has been mechanically inputted into a computer via a text file. Examples of this latter type of system are word processors and photo-typesetters.

These conventional systems have limitations which decrease the efficiency of processing information from a hard copy document. For example, the systems discussed above are limited in their application to MICR, OCR, or typewritten data. Parsing and processing data is limited to the particular requirements of the particular computer application which requires the input data. In addition, in these conventional systems, the actual hard copy document must be retained for future reference at great expense.

In a sophisticated computer network, different users may require different portions of the information contained on a hard copy document. For example, if the hard copy document is an invoice returned with payment of a bill, the accounting department may need all of the monetary information contained on the bill while the mailroom may need only customer address information, to update a customer's address. Therefore, there is a need for a system in which specific information from a hard copy document can be selectively distributed to various users.

Another problem with conventional systems is that users, even within the same company, may require that the information extracted from a hard copy document be transmitted to a particular application program in a specific transmission format. For example, one department in a company may use a particular application program which must receive information using a particular character as a delimiter and other departments may require the information in a different format using different delimiters.

Another problem, particularly for small businesses, is that current systems can not efficiently accommodate the inputting of information from a diversity of hard copy documents. A large business which receives many forms in the same format can afford a system which inputs a high volume of information in that format into memory. For example, it is

2

cost-effective for a bank which processes hundreds of thousands of checks a month to buy a dedicated machine which can read information off of checks having a rigidly defined, or fixed, format. However, as the diversity of forms received by a business increases relative to the number of forms that must be processed, it becomes less cost-effective to design a dedicated machine for processing each type of form format. This problem is particularly significant in small businesses which may, for example, receive fifty invoices a month, all in different, non-fixed, formats. It is frequently not cost-effective for a small business to design dedicated systems for inputting information in each of these various formats. This leaves a small business with no other practical alternative than to manually input the information off of each invoice each month.

SUMMARY OF THE INVENTION

It is an object of the invention, therefore, to provide an application program interface which allows a user to select specific portions of information extracted from a diversity of hard copy documents and allows the user to direct portions of this information to several different users in accordance with the needs of the particular user.

It is also an object of the invention to provide a cost-effective system for inputting hard copy documents which can accommodate hard copy documents in a diversity of formats.

It is another object of the invention to provide an application program interface which allows a user to put information, which is to be transmitted, into a particular transmission format, based upon the needs of the receiver of the information.

It is a further object of the invention to provide an application program interface which will allow the extraction, selection, formatting, routing, and storage of information from a hard copy document in a comprehensive manner such that the hard copy document itself need not be retained.

It is another object of the invention to provide a system which reduces the amount of manual labor required to process information originating from a hard copy document.

A further object of the invention is to reduce the time required to process information originating from a hard copy document so that a higher volume of transactions involving hard copy documents can be processed.

The invention provides an application program interface which inputs a diversity of hard copy documents using an automated digitizing unit and which stores information from the hard copy documents in a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which define portions of the stored document information required by a particular application unit. Selected stored document information is then formatted into the transmission format used by the particular application program based on transmission format instructions. The transmission formatted selected stored document information is then transmitted to the particular application program. The hard copy documents may contain textual information or image information or both.

The interface operates in three different modes.

In a first mode, the interface extracts all of the information from hard copy documents and stores this information in memory. Parsing of various portions of the extracted information is performed in accordance with content instructions.

In a second mode, the user operates interactively with the interface by use of a display and an input device, such as a mouse. In this second mode, a hard copy document is inputted and displayed on the display. The interface then prompts the user to identify the location of various information. For example, the interface can ask the user to identify the location of address information on the hard copy document. In response, the user positions the mouse to identify address information using a cursor. The identified information is then stored as address information in memory. Subsequently, the interface again prompts the user to identify other pieces of information, which are then stored in the appropriate locations in memory. This process proceeds until all of the information which is desired to be extracted off of the hard copy document is stored in memory.

In a third mode of operation, selected portions of information are extracted off of hard copy documents in accordance with predetermined location information which has been specified by the user. For example, the user can define a template which specifies the location of information on hard copy documents. Templates can be formed in conjunction with second mode operation. Alternatively, the user can instruct the interface to search hard copy documents for a particular character or symbol, located on the hard copy documents. The information desired to be extracted off of the hard copy documents is specified relative to the location of this character or symbol.

The interface can also prompt or receive from an applications program or another information processing system, required information, content instructions, and format instructions.

Other objects, features, and advantages of the invention will be apparent from the following detailed description of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be described in further detail below with reference to the accompanying drawings, in which:

FIG. 1 illustrates hardware for implementing a preferred embodiment of the instant invention;

FIG. 2 illustrates an example of a hard copy document containing information to be processed by the instant invention;

FIGS. 3A and 3B are enlarged views of the computer of FIG. 1 used to explain how the invention interactively prompts a user to identify information;

FIG. 4 is an overall data flow diagram for the FIG. 1 preferred embodiment;

FIG. 5 is a detailed input data flow diagram for the FIG. 1 preferred embodiment;

FIG. 6 is a detailed information processing data flow diagram for the FIG. 1 preferred embodiment;

FIG. 7 is a more detailed information processing data flow diagram for the maintain library module of FIG. 6;

FIG. 8 is a more detailed information processing data flow diagram for the maintain definitions module of FIG. 6;

FIG. 9 is a more detailed information processing data flow diagram for the process document module of FIG. 6;

FIG. 10 is a detailed output data flow diagram for the FIG. 1 preferred embodiment;

FIG. 11 lists data corresponding to the hard copy document of FIG. 2;

FIGS. 12A, 12B, and 12C illustrate examples of data which can be selected from the extracted data of FIG. 11 in accordance with content instructions;

FIGS. 13A, 13B, and 13C illustrate examples of the data of FIGS. 12A, 12B, and 12C formatted in accordance with various transmission format instructions to form input files; and

FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Hardware

The invention provides an interface between information originating from a hard copy document and a computer application unit which uses the information. The computer application unit can be a particular computer application program or a device which is controlled in accordance with instructions or information from the hard copy document. The invention also allows storing a copy of the hard copy document in a memory and retrieving the copy of the hard copy document. By providing a comprehensive and integrated system which can accommodate almost all of the possible uses of information contained on a hard copy document, the instant invention allows for a paperless office.

The invention includes hardware and software necessary to extract, retrieve, and process information from the hard copy document. A copy of the actual image of the hard copy document is stored in memory. Textual information extracted from the hard copy document is also stored in memory. Textual information is information, such as alphanumeric characters, which is recognized on the hard copy document and which is stored in a form which corresponds to the particular recognized character. For example, the extracted characters can be stored in the ASCII format in an electronic memory.

The user can have all of the information extracted from the hard copy document and stored in memory. Alternatively, the interface can interactively prompt the user to identify specific pieces of information for storage. The interface can also extract specific pieces of information using a predefined template. The interface can also prompt or receive from another information processing system or an applications program desired information, content instructions, and format instructions.

The instant invention also provides for parsing information extracted from the hard copy document and for directing this parsed information to specific users or application programs as an input file.

The invention also permits the user to define the transmission format of the input file for a particular computer application unit.

FIG. 1 illustrates hardware for implementing a preferred embodiment of a hard copy document to application program interface according to the instant invention. The interface 200 processes information extracted off of hard copy document 100 and provides information to application units 270 in a form required by each particular application unit. The interface extracts information off of a hard copy document 100 utilizing a scanner 210. The scanner 210 can be any type of scanner which extracts information off of hard copy documents, for example, an Optical Reader.

The scanned information is stored in a scanner memory 220 or in main memory 250, as will be described in greater detail below. If main memory 250 or another memory is available to store the scanned information, then scanner memory 220 can be omitted.

The information from scanner memory 220 or main memory 250 is transmitted to computer 230. In the preferred embodiment, computer 230 includes a display 232, a keyboard 234, and a mouse 236. The display 232 displays an image of the hard copy document itself and/or information necessary to process the information extracted off of the hard copy document.

The computer 230 is used to select portions of the stored document information contained in memory in accordance with content instructions which define portions of the stored document information required by an application unit. These content instructions may be provided by the application program. Alternatively, the content instructions can be inputted via an input device such as a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

The computer 230 is also used to format selected stored document information into the transmission format used by an application unit based on transmission format instructions. The transmission format instructions may be provided by the application program. Alternatively, the transmission format instructions can be inputted via a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

Thus, the computer 230 is used to generate an input file for a particular application unit. The computer 230 is connected to scanner memory 220, main, or permanent, memory 250, a printer 260, and application units 270, via a bus 240. Although FIG. 1 illustrates use of a bus to connect components together, it is understood that any routing or connecting link, implemented in hardware or software or both, can be employed instead of, or in addition to, a bus. Instructions to or in the computer 230 control the main memory 250, the printer 260, the application units 270, and the bus 240. Instructions to or in computer 230 can also control exchanges of information with scanner memory 220.

When the computer 230 generates an input file for a particular document, the computer 230 can send this input file directly to an application unit or can store this input file in the main memory 250 until required by an application unit. The main memory 250 may also optionally store a copy of the image information for the hard copy document and the textual information for the hard copy document. Thus, the image information and textual information from the hard copy document can be retrieved and printed out on printer 260. In addition, image and textual information stored in scanner memory 220 or in main memory 250 can be used to form additional input files at the time of input or at a later time, based on content instructions and transmission format instructions. Thus, the invention can, at the discretion of the user, eliminate the need to retain copies of hard copy documents, permitting a paperless office.

The application units 270 include particular application programs and devices which are controlled in accordance with information contained on hard copy document 100.

FIG. 2 illustrates an example of a hard copy document 100 which contains information to be processed by the instant invention. The document illustrated in FIG. 2 is a bill from XYZ Corporation to customer ABC Corporation. FIG. 2 is only an example of a type of document that can be processed by the instant invention.

In a first operational mode, the scanner 210 stores all of the information extracted of f of hard copy document 100 in the scanner memory 220 or, alternatively, in main memory 250. The extracted information is stored in two forms. The actual image of the hard copy document 100 is stored as image information in the scanner memory 220. In addition,

the scanner memory 220 stores textual information recognized on the hard copy document 100 by, for example, employing standard character recognition software. In the preferred embodiment, the textual information is stored in ASCII format. The scanner memory 220 can be, for example, an electronic, magnetic, or optical memory.

FIG. 3A illustrates an enlarged view of the computer 230 of FIG. 1. This view will be used to describe a second mode of operation. In this second mode of operation, the hard copy document 100 is scanned and a copy of the document 100 is displayed on display 232 of computer 230, based on the contents of information temporarily stored in scanner memory 220. After the document is displayed on display 232, the computer 230 interactively prompts the user to identify the location of specific pieces of information on the hard copy document. In the FIG. 3A illustration, this prompt message is indicated as the message beginning with the arrow.

For example, the prompt message can ask the user to identify the location of account number information on the hard copy document. The user then uses an input device, such as keyboard 234 or mouse 236 or a touch screen, notepad, voice recognition device, or other input device to position a cursor on the display to identify the location of the information requested by the prompt message. For example, the cursor could be used to define a block (which could be highlighted) containing the requested information, followed by a mouse "enter" click. In this example, the user would move the mouse to identify the location of the account number information contained on the hard copy document 100. The computer 230 then stores the information which has been identified by the user as account number information in the appropriate address or subfile or as the appropriate variable or parameter, or data filed, in memory. The computer then prompts the user to identify the location of other information on the hard copy document, such as, statement date information. The process proceeds until all of the desired information has been stored into the appropriate locations in memory.

FIG. 3B illustrates a variation of the second mode for interactively prompting the user for information. In FIG. 3B, the display is split into two portions. A left-hand portion 232L displays the image of the hard copy document and a right-hand portion 232R displays the required application program information. For example, in FIG. 3B, portion 232R displays a spreadsheet used by an application program. While observing the split display, the user can input instructions to associate specific pieces of information on the hard copy document (for example, the vendor name indicated by the mouse arrow 232A) with particular subfiles in memory (for example, the vendor field next to which the cursor 232C appears), using a mouse or other input device(s) or both. The split display also allows the user to generate content format instructions while observing the information required for a particular application program on the right-hand portion.

These second modes of operation are efficient for small businesses which receive a small number of a wide variety of invoices, since the user does not necessarily have to store all of the information that appears on the hard copy document. A further advantage is that data input is quicker, easier, and more accurate than with previous keyboard methodology. In addition, by specifying the location on the hard copy document of information, the user may optionally create a template, to be described in further detail below, for each different type of invoice. This template is stored for future use when another hard copy document in the same format is received.

More specifically, instructions from computer 230 can direct the scanner 210 and scanner memory 220, and/or main memory 250, to scan and/or store only specific portions of hard copy document 100. After the interactive prompts required to obtain information for a desired application program, the unused information stored in scanner memory 220 or 250 can be erased. Further, scanning of a second identical document can be limited to only those portions of the document which contain needed information.

More specifically, in FIG. 2, the lines 10 drawn around certain portions of the document represent the areas which the user has previously identified as the portions of a document to be extracted by the scanner 210 and stored in scanner memory 220 and/or main memory 250. Since the logo 20 and the message 30 have not been identified as an area to be scanned and stored, these areas are not scanned and stored in subsequent documents. Since the user has previously associated each of the areas 10 with a specific subfile of information, e.g., the account number, the scanned information is stored in memory locations corresponding to that subfile.

Data Processing

FIGS. 4-10 illustrate the flow of data in the FIG. 1 preferred embodiment. FIG. 4 illustrates the overall data flow for the FIG. 1 preferred embodiment. The preferred embodiment includes an input process module 1.0, an information processing module 2.0, and an output processing module 3.0. The information processing module 2.0 is equipped to receive instructions from and transmit information to a user. The information processing module 2.0 can also transmit to and receive information from a remote external device through communication interface 4.0. Input process module 1.0 and output processing module 3.0 can also access communication interface 4.0. A module is implemented in hardware, software, or a combination of hardware and software. The specific implementation for a particular business application depends upon a variety of factors, for example, the relative costs of hardware and software implemented systems, the frequency with which a user will want to expand or modify the system, and the like.

FIG. 5 is a more detailed diagram of the input process module 1.0 of FIG. 4. The input process module 1.0 includes a character input module 1.1, an image input module 1.2, and, in the preferred embodiment, a character recognition device 1.3. The character input module inputs textual information, such as alphanumeric characters, from an input device such as keyboard 234. The image input module 1.2 inputs image information, for example, a digitized image of the actual appearance of hard copy document 100. Textual information can include textual input from an input device such as keyboard 234 and textual information extracted from the document by character recognition device 1.3. Both types of information comprise an input document which is transmitted to information processing module 2.0. In the FIG. 1 preferred embodiment, the processing performed by input process module 1.0 occurs in scanner memory 220, computer 230, and main memory 250.

FIG. 6 illustrates information processing data flow for the FIG. 1 preferred embodiment, that is, FIG. 6 illustrates data flow in the information processing module 2.0.

The information processing module 2.0 includes a maintain library module 2.1, to be described in further detail below in conjunction with FIG. 7, a maintain definitions module 2.2, to be described in further detail below in conjunction with FIG. 8, and a process document module 2.3 to be described in further detail below in conjunction with FIG. 9.

The information processing module 2.0 is the module which coordinates and drives the entire system. In the preferred embodiment, the information processing module 2.0 is implemented primarily by computer 230.

FIG. 7 illustrates information processing data flow in the maintain library module 2.1. The maintain library module 2.1 maintains a library of image information, for example, a digitized image representing the actual appearance of the hard copy documents for reference during processing. This library can be incorporated within scanner memory 220, main memory 250, or another independent memory, for example, a RAM disk. The maintain library module 2.1 includes a store document module 2.1.1, a correct errors module 2.1.2, a retrieve document module 2.1.3, and a document file 2.1.4. These modules operate collectively to store, retrieve, and correct document information.

The store document module 2.1.1, prior to routing the document to the document file 2.1.4, may provide information on recognition errors which may have occurred while inputting the document. For example, the store document module 2.1.1 identifies that a character contained on hard copy document 100 was not recognized. The store document module 2.1.1 also optionally causes a copy of the document and its parsing to be displayed on the display 232 for confirmation by the user. The user may utilize this opportunity to identify any errors in the displayed document and, in conjunction with the correct errors module 2.1.2, to revise the document's parsing, if necessary, prior to storage of the document in memory. The module 2.1.1 also provides a facility for the user to name a particular hard copy document for cataloging, storage, and retrieval purposes. After the document is named, the store document module 2.1.1 stores copies of the document in the document file 2.1.4.

The correct errors module 2.1.2 processes instructions from the user to correct errors identified by the store document module 2.1.1 and errors that have been spotted by the user during the confirmation process.

The retrieve document module 2.1.3 permits the user to retrieve a copy of a document previously stored in the document file 2.1.4. As described above, long-term storage is provided by main memory 250, if necessary.

FIG. 8 illustrates a more detailed information processing data flow diagram for the maintain definitions module 2.2 of FIG. 6. The maintain definitions module 2.2 allows the user to define system and document parameters and maintains the definitions of these system and document parameters. The maintain definitions module 2.2 includes a define template module 2.2.1 which allows the user to specify the location of information on the document. This information provided by the user defines a template which is used to extract information off the document and to associate the extracted information with a particular variable or subfile. These templates are illustrated by boxes 10 in the FIG. 2 example of a hard copy document. The maintain definitions module 2.2 can also access templates previously defined by the user and stored in main memory 250. Templates can also be provided as part of software packages developed by program developers.

The maintain definitions module 2.2 also includes a define relationships module 2.2.2. The define relationships module 2.2.2 allows the user to define data relationships, or logical relationships, between pieces of information extracted from the hard copy document. These pieces of information are then used to generate an input file for a selected computer application unit. The user defines these relationships by

content instructions. Alternatively, content instructions to define relationships can be provided by application software. If the user provides these content instructions, the content instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. Examples of content instructions, data, and logical relationships will be described in further detail in conjunction with FIGS. 11 and 12A, 12B, and 12C.

The maintain definitions module 2.2 also includes a define format module 2.2.3. The define format module 2.2.3 allows the user to define transmission formats for an input file which is then transmitted to a selected computer application unit. Selection of the transmission format of the input file is accomplished by the user through use of transmission format instructions. Alternatively, the applications software itself can generate its own transmission format instructions. When the user must specify transmission format instructions, the transmission format instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. A further description of various transmission formats will be provided below in conjunction with FIGS. 12A, 12B, 12C, 13A, 13B, and 13C.

A select definitions module 2.2.4 is also included in the maintain definitions module 2.2. The select definitions module 2.2.4 allows the user to store and select a set of definitions to be used for processing the document. The definitions identify pieces of information on the document by, for example, absolute location, variable location, or relative location, or by proximity to key words and/or symbols. These definitions are described in further detail below by way of an illustrative example.

FIG. 9 illustrates a more detailed information processing data flow diagram for the process document module 2.3. The process document module 2.3 processes the document after the document has been stored in the system. The process document module 2.3 gathers the appropriate information which has been stored, and creates input file(s) 2.3.3 for the selected application unit. The process document module 2.3 then transmits the input file(s) via bus 240 and/or communication interface 4.0 to an application unit 270, an output device such as printer 260, or to main memory 250.

The process document module 2.3 includes an extract data module 2.3.1. This module extracts data off of the document in accordance with the user's instructions, for example, the user-defined template, or through the interactive mode.

The process document module 2.3 also includes a preapplication process module 2.3.2 which gathers and associates information extracted from the document in accordance with content instructions. This module prompts the user for any additional information required to satisfy the relationships defined by the content instructions. The preapplication process module 2.3.2 also places the selected information into the Transmission format defined by the transmission format instructions.

The preapplication process module 2.3.2 also generates the input file 2.3.3 for the selected application in accordance with the appropriate instructions. The input file 2.3.3 is then transmitted to bus 240 and/or communication interface 4.0 for transmission to a particular application unit 270.

FIG. 10 illustrates a detailed output data flow diagram for output module 3.0. Output module 3.0 outputs a textual and/or image copy of the document. In the FIG. 1 preferred embodiment, output module 3.0 is implemented by printer 260, associated software, and associated interface circuitry.

Operation

Examples of operation of a preferred embodiment will now be described.

The user enters the system by providing instructions to the information processing module 2.0. The user then instructs the information processing module 2.0 to conduct maintain library processing, maintain definitions processing, or process document processing.

If the user selects maintain library processing, the user then provides instructions to maintain or modify the document library through the maintain library module 2.1. For example, the user can direct the inputting and storage of a hard copy document 100 or can retrieve and output a document. The user requests inputting of a document through the store document module 2.1.1. The system then prompts the user to specify a storage location for the inputted document. The document is then read-in by the input process module 1.0. A textual copy and/or an image copy are stored into the document file 2.1.4. Errors which have occurred during inputting are identified and corrected by the correct errors module 2.1.2 and the user. The corrections are reflected in the document information stored in document file 2.1.4.

The retrieve document module 2.1.3 is used to retrieve and output a document. The system prompts the user to specify the storage location of a document and the type of document copy, for example, a textual or an image copy, to be outputted. The document is then outputted by the output process module 3.0.

If the user initially selected maintain definitions processing, the user would instruct the system to maintain and/or modify parameter definitions through the maintain definitions module 2.2. For example, the user can define and maintain a document template for extracting selected portions of information of f of the hard copy document. The user can use the template to extract selected portions of information off of the hard copy document when the document is originally inputted, or alternatively, the user can use the template to identify selected portions of information for extraction off of an image copy of the document. In creating the template, the user identifies pieces of information on the document to be extracted and assigns a variable name, or subfile, to each piece of data.

The location of data to be extracted can be defined in a number of ways other than by use of a template. For example, the user can designate the absolute location of information on the document with respect to a grid overlaid on the document, e.g., always on line 3, starting, in column 1. The user can also identify information by specifying the relative location of information to be extracted, e.g., always two lines below the piece of data named "salutation", starting in column 3. The user can also specify the location of information to be extracted by variable location specification. For example, if the hard copy document is a letter, the module would conduct a key word search for the term "Dear Sir:". Wherever this term "Dear Sir:" is located, this piece of data would be associated with the variable specified by the user, for example, the variable "salutation." In addition, a defined set of conventional symbols can be used to signify certain recurring data items for the convenience of users of the instant invention. For example, a "@" symbol can be used to delineate the vendor name as follows: "@XYZ Corporation@". Other examples of the use of symbols to delineate information will be described with reference to FIG. 14.

The maintain definition module 2.2 is also used to maintain data relationships in accordance with content instruc-

tions and to maintain input file formats in accordance with transmission format instructions. Relationships are defined and maintained between pieces of data, specified by, for example, the names of variables, through the define relationships module 2.2.2. The names of pieces of data on the document are retrieved by, for example, the define template module 2.2.1, and are passed to the define relationships module 2.2.2. The user may then provide any additional pieces of data needed to generate an input file for a particular application program or unit, such as an input file line number. The user, the applications software, and/or instructions previously stored in memory then establishes the contents of the input file by defining relationships between pieces of data using content instructions. Specific examples of content instructions will be discussed below in conjunction with FIGS. 11, 12A, 12B, 12C, 13A, 13B, and 13C.

The user and/or the applications software defines and maintains the transmission format of the input file to be used by a particular application program or unit through the define format module 2.2.3 in accordance with transmission format instructions. This is accomplished by defining the parameters to be used by the preapplication process module 2.3.2 in generating an input file. Parameters which would typically be required to generate an input file would include the character type, e.g., text or pixel; delimiters used between pieces of data, e.g., a slash or a semicolon; end of line characters, e.g., a carriage return or a line feed; and end of file characters. Examples of transmission formats will be described in further detail below in conjunction with FIGS. 11, 12A, 12B, 12C, 13A, 13B, and 13C.

If the user initially selected process document processing, the interface will then proceed to process the document through use of the process document module 2.3. For example, the user can extract specific portions of data from an image copy of a document, can generate an input file for transmission to an application program, or can directly process information interactively with an application program.

If the user desires to extract specific portions of data from an image copy of a hard copy document which has already been stored in memory, the user uses the extract data module 2.3.1 to identify a document to be processed. The document is then retrieved by the retrieve document module 2.1.3 and passed to the extract data module 2.3.1. The user can also select parameter definitions through the select definitions module 2.2.4.

The selected document template or parameter definition is passed to the extract data module 2.3.1. The extract data module 2.3.1 extracts pieces of data from the image copy of the document, as defined by the document template definition or the parameter definitions or both. This document data is then passed to preapplication process module 2.3.2.

The interface generates input file(s) 2.3.3 by use of the preapplication process module 2.3.2. The selected data relationship definition, as defined by the content instructions, and the selected record format definitions, as defined by the transmission format instructions, are passed to the preapplication process module 2.3.2. The preapplication process module 2.3.2 assembles the input file in accordance with the content instructions. The preapplication process module 2.3.2 also prompts the user for any additional pieces of data which need to be provided by the user. The input file is converted to the desired transmission format in accordance with the transmission format instructions. This physically formatted data is then stored in the input file 2.3.3.

The user can also use an application program to process information by loading the particular application program

into the computer 230 rather than by sending the input file to a remote application unit 270.

An illustrative example of the processing described above will now be described.

5 The user inputs instructions via keyboard 234 or another input device which indicate that the user desires to input and store a document. The computer 230 then prompts the user for the name of the document. In this example, the user desires to input the document of FIG. 2 and therefore names 10 the document "XYZ Corp. Bill Dec. 1, 1986." The computer then prompts the user to feed the hard copy document 100 into the scanner 210. The image of the hard copy document is displayed on display 232. The computer then prompts the user to identify the account number on the document. By use 15 of the mouse 236 or other input device to position a cursor on the display, the user indicates the location of the account number. The account number is then read-in to a subfile named "Account Number." This process proceeds until all of the desired information has been read-in and stored.

20 In this particular example, no errors were encountered while inputting the document. The user then directs that the document be stored for future reference in a document file.

Some time later, the user desires to retrieve and output the document and to generate input files based on information from the document. The computer 230 prompts the user for the name of the document and the type of output. The user responds with "XYZ Corp. Bill Dec. 1, 1986" for a printed textual copy. The document is then retrieved from the document file and passed to the printer 260 for printing.

25 In order to generate an input file for a specific application program, the user selects the option to define a document template for use when each month's XYZ Corporation bill arrives. Accordingly, the user instructs the system to display a copy of an XYZ Corporation bill on the display 232. The user then identifies pieces of data by absolute locations. That is, the user assigns specific names to information located at specific portions of the document. In this example, the user 30 will input the following information:

40 Vendor-text, line 1, one line, column 1, 80 characters;
Account number-numeric, line 6, one line, column 25, 9 characters;

45 Statement date-date, line 9, one line, column 25, 8 characters;
Payment date-date, line 11, one line, column 25, 8 characters;

50 Previous balance-currency, line 7, one line, column 75, 9 characters;
New charges-currency, line 8, one line, column 75, 9 characters;

55 Other debits-currency, line 10, one line, column 75, 9 characters;
Finance charges-currency, line 12, one line, column 75, 9 characters;

60 Payments-currency, line 13, one line, column 75, 9 characters;
Other credits-currency, line 14, one line, column 75, 9 characters;

65 New balance-currency, line 15, one line, column 75, 9 characters.

The user also identifies data with variable locations. In this particular example, a variable location is specified as follows:

65 Heading 2-line, value ="Mail To."
The identification of Heading 2 as line information means that the system will search for occurrences of the character

13

string "Mail To:" and assign the line number which contains this character string to Heading 2.

The user also identifies data by relative locations. In this example, the user identifies the following relative location:

Mail To-text, Heading 2+1, 3 lines, column 60, 25 characters per line.

The instructions above instruct the system to assign the textual information beginning on one line after Heading 2 and continuing for 3 lines, in column 60, to the Mail To subfile.

As an alternative to inputting the actual line, column, and character numbers, the user can identify desired portions of the document by blocking, or highlighting, the desired portions using the mouse or other input device. In this case, the computer converts the highlighted portions into corresponding line, column, and character numbers.

FIG. 11 lists data corresponding to the hard copy document of FIG. 2 and the associated variable or subfile names.

Next, the user desires to define data relationships in accordance with content instructions. Examples of the type of contents which can be specified by a user are illustrated in FIGS. 12A, 12B, and 12C.

In this particular example, three separate departments of ABC Corporation require information from the XYZ Corporation bill. The first department requires vendor, account number, statement date, payment date, previous balance, new charges, debits, finance charges, payments, and new balance information. The second and third departments require mail to information and previous balance information. Each of these departments have their own application program which utilizes this information.

The user employs content instructions to designate how pieces of information, which have been extracted off of hard copy document 100, are directed to particular departments, that is, particular application programs. FIG. 12A illustrates the contents of the information to be transmitted to the first department. FIG. 12B illustrates the information to be transmitted to the second department. FIG. 12C illustrates the information to be transmitted to the third department. The content instructions, therefore, parse the information shown in FIG. 11 to various application programs, as shown by FIGS. 12A, 12B, and 12C. Content instructions can also be used to identify additional pieces of data which are required for the input files of the particular application programs. In this particular example, the specific application programs from the three departments all require numeric record number information, numeric horizontal position information, numeric vertical position information, and date received information. The horizontal and vertical position information is used by the application program to specify the location of the received information on a spreadsheet application program, in this example. The user may know in advance the content format required by each application program, that is, in this example, the location and type of information specified on the spreadsheet. The user may also employ the split display mode described with reference to FIG. 3B to generate content format instructions.

Using the content instructions, the user establishes the following contents for the input file corresponding to FIG. 12A:

Record number, horizontal position, vertical position, vendor;

Record number, horizontal position, vertical position, account number;

Record number, horizontal position, vertical position, statement date;

Record number, horizontal position, vertical position, date received;

14

Record number, horizontal position, vertical position, payment date;

Record number, horizontal position, vertical position, previous balance;

Record number, horizontal position, vertical position, new charges;

Record number, horizontal position, vertical position, finance charges;

Record number, horizontal position, vertical position, payments;

Record number, horizontal position, vertical position, new balance.

Next, transmission format instructions are employed to define the transmission format of the input file for a specific application program or unit. FIG. 13A illustrates the transmission input file corresponding to FIG. 12A. FIG. 13B illustrates the transmission input file corresponding to FIG. 12B. FIG. 13C illustrates the transmission input file corresponding to FIG. 12C. A comparison of FIGS. 12B and 12C reveals that FIGS. 12B and 12C have the same contents. However, the information illustrated in FIG. 12B is being sent to a different application program than the information in FIG. 12C. These application programs require different transmission input formats as illustrated in FIGS. 13B and 13C. More specifically, the application program that receives the input file illustrated in FIG. 13B uses the greater than sign as a delimiter whereas the application program which receives the transmission input file shown in FIG. 13C uses a back-slash as the delimiter.

After the contents and the transmission format for the input file have been defined, and any additional information has been inputted, the input file is assembled and transmitted to the particular application program.

FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention. The hard copy document illustrated in FIG. 14 is first scanned and information from the hard copy document is stored into a memory. The interface 200 then identifies portions of the hard copy document corresponding to various variables by recognizing a defined set of symbols. In the FIG. 14 example, triangles delineate the mailing address, circles delineate the statement date, and squares delineate the new charges. Information from these portions of the hard copy document is stored in the corresponding memory locations or subfiles for each variable. The same set of symbols can be used to identify the same information from one document to the next. Thus, even if the physical formats of documents are not fixed from one document to the next, a diversity of hard copy documents can be processed without manually inputting data by recognition of the defined symbols.

Examples of readily available application programs are Quicken and Lotus 1,2,3 both of which are widely utilized in the business community. Quicken, for example is an easy-to-utilize program for writing checks and preparing business records. Payee, amount and address information may readily be transmitted from scanner memory 220 and/or main memory 250 to the Quicken application program for check writing functions and ledger keeping purposes. Lotus is a well known spreadsheet program which may process data input into specified cells once this data is placed in conventional Lotus format.

Thus, the instant invention provides an integrated and comprehensive system for handling information from a hard copy document, thus permitting a paperless office. In addition, the invention permits data, extracted from a hard

copy document, to be easily manipulated into various logical and transmission formats required by a particular application unit. The invention also provides a low cost system for inputting information from a wide variety of hard copy documents into a memory.

The foregoing description has been set forth merely to illustrate preferred embodiments of the invention and is not intended to be limiting. Modifications are possible without departing from the scope of the invention.

For example, letters, checks, forms, pictures, reports, music scores, film, and other types of hard copy documents can be processed by the invention for accounts payable/receivable accounting, inventory control, record keeping, budgeting, data base management, music transcription, forms processing, computerized art, survey and questionnaire processing, statistical data analysis, correspondence processing and other applications.

Other automated digitizing units can be used in addition to or as an alternative to use of the scanner 210 as an input unit. Any electrical, magnetic, or optical device which extracts information off of a hard copy document, thereby eliminating the need to manually input significant amounts of information from the hard copy document is suitable for use as an automated digitizing unit. In addition, information can be input by user responses and digital and analog signals generated from various devices, and from computer files from other computer systems. Suitable hardware for inputting data includes a keyboard, a light pen, a mouse, a touch screen, a laser scanner, a microphone, a tablet, a disk drive, a magnetic tape drive, and a modem.

The interface 200 can also output information in forms other than a hard copy of textual or image information. For example, the interface 200 can output system responses, computer files, and digital and analog signals for transmission to other computer systems or to control systems. Suitable hardware for outputting information includes a disk drive, a magnetic tape drive, a cathode ray tube, a plasma screen, a printer, a plotter, a film developer, an amplifier, and a modem.

Since modifications of the described embodiments incorporating the spirit and substance of the invention may occur to persons skilled in the art, the scope of the invention should be limited solely with respect to the appended claims and equivalents.

What is claimed is:

1. A multimode information processing system for inputting information from a document or file on a computer into at least one application program according to transmission format instructions, and to operate in at least one of:
 - a. a definition mode wherein content instructions, at least one of which is not a location of said information derived from a pre-scanned image of a blank form, are used to define input information from within said document or file required by said at least one application program; and
 - b. an extraction mode to parse at least a portion of said document or file to automatically extract at least one field of information required by said at least one application program and to transfer said at least one field of information to said at least one application program.

2. A multimode information processing system as recited in claim 1, wherein said definition mode is operative to store said content instructions in said system in association with said extraction mode.

3. A multimode information processing system as recited in claim 1, wherein said definition mode is operative to store

said content instructions in said system as a template for use in said extraction mode.

4. A multimode information processing system as recited in claim 1, wherein said extraction mode is operative to match at least a portion of said document or file with a template created with or available to said system.

5. A multimode information processing system as recited in claim 2, wherein said extraction mode is operative to match at least a portion of said document or file with a template created with or available to said system.

10 6. A multimode information processing system as recited in claim 1 in which said document or file is on a first computer, and said application program is on a second computer.

15 7. A multimode information processing system as recited in claim 3 in which said document or file is on a first computer, and said application program is on a second computer.

20 8. A multimode information processing system as recited in claim 7 in which said template is stored in said first computer.

9. A multimode information processing system as recited in claim 7 in which said template is stored in said second computer.

25 10. A multimode information processing system as recited in claim 4 in which said document or file is on a first computer, and said application program is on a second computer.

11. A multimode information processing system as recited in claim 10 in which said template is stored in said first computer.

30 12. A multimode information processing system as recited in claim 10 in which said template is stored in said second computer.

13. A multimode information processing system as recited in claim 5 in which said document or file is on a first computer, and said application program is on a second computer.

35 14. A multimode information processing system as recited in claim 13 in which said template is stored in said first computer.

40 15. A multimode information processing system as recited in claim 13 in which said template is stored in said second computer.

16. A multimode information system as recited in claim 3, 45 4, or 5 in which said document or file, said at least one application program and/or said template is/are distributed among more than one computer.

17. A multimode information processing system for inputting information from a document or file on a computer into at least one application program according to transmission format instructions, and to operate in:

- a. a definition mode wherein content instructions, at least one of which is not a location of said information derived from a pre-scanned image of a blank form, are used to define input information from within said document or file required by said at least one application program; and
- b. an extraction mode to parse at least a portion of said document or file to automatically extract at least one field of information required by said at least one application program and to transfer said at least one field of information to said at least one application program.

18. A multimode information processing system as recited in claim 17, wherein said definition mode is operative to store said content instructions in said system in association with said extraction mode.

19. A multimode information processing system as recited in claim **17**, wherein said definition mode is operative to store said content instructions in said system as a template for use in said extraction mode.

20. A multimode information processing system as recited in claim **17** wherein said extraction mode is operative to match at least a portion of said document or file with a template created with or available to said system.

21. A multimode information processing system as recited in claim **18**, wherein said extraction mode is operative to match at least a portion of said document or file with a template created with or available to said system.

22. A multimode information processing system as recited in claim **17** in which said document or file is on a first computer, and said application program is on a second computer.

23. A multimode information processing system as recited in claim **19** in which said document or file is on a first computer, and said application program is on a second computer.

24. A multimode information processing system as recited in claim **23** in which said template is stored in said first computer.

25. A multimode information processing system as recited in claim **23** in which said template is stored in said second computer.

26. A multimode information processing system as recited in claim **20** in which said document or file is on a first computer, and said application program is on a second computer.

27. A multimode information processing system as recited in claim **26** in which said template is stored in said first computer.

28. A multimode information processing system as recited in claim **26** in which said template is stored in said second computer.

29. A multimode information processing system as recited in claim **21** in which said document or file is on a first computer, and said application program is on a second computer.

30. A multimode information processing system as recited in claim **29** in which said template is stored in said first computer.

31. A multimode information processing system as recited in claim **29** in which said template is stored in said second computer.

32. A multimode information system as recited in claim **19, 20, or 21** in which said document or file, said at least one application program and/or said template is/are distributed among more than one computer.

33. An application program interface comprising:

- a. utilities for enabling selection of portions of a stored document or file in accordance with content instructions, at least one of which is not a location of said information derived from a pre-scanned image of a blank form, as selected stored document information, said content instructions designating portions of said stored document or file required by an application unit;
- b. utilities for enabling formatting of said selected stored document information into a transmission format used by said application unit based on transmission format instructions; and
- c. utilities for automatically enabling transmission of formatted selected stored document information to said application unit in an extraction mode.

34. An application program interface as recited in claim **33** further comprising utilities for enabling storage of said content instructions.

35. An application program interface as recited in claim **33** or **34** further comprising:

utilities for enabling definition of a template which associates portions of said document or file with specific variables.

36. An application program interface as recited in claim **33** or **34** which operates in a distributed computing environment.

37. An application program interface as recited in claim **35** which operates in a distributed computing environment.

38. An method of doing business comprising the steps of: inputting information from a document or file on a computer into at least one application program according to transmission format instructions; processing the information in at least one of:

a. a definition mode wherein content instructions, at least one of which is not a location of said information derived from a pre-scanned image of a blank form, are used to define input information from within said document or file required by said at least one application program; and

b. an extraction mode to parse at least a portion of said document or file to automatically extract at least one field of information required by said at least one application program and to transfer said at least one field of information to said at least one application program.

39. A multimode information processing system as recited in claim **30** wherein said definition mode is operative to store said content instructions in said system in association with said extraction mode.

40. A multimode information processing system as recited in claim **38**, wherein said definition mode is operative to store said content instructions in said system as a template for use in said extraction mode.

41. A multimode information processing system as recited in claim **38**, wherein said extraction mode is operative to match at least a portion of said document or file with a template created with or available to said system.

42. A multimode information processing system as recited in claim **39**, therein said extraction mode is operative to match at least a portion of said document or file with a template created with or available to said system.

43. A multimode information processing system as recited in claim **38** in which said document or file is on a first computer, and said application program is on a second computer.

44. A multimode information processing system as recited in claim **40** in which said document or file is on a first computer, and said application program is on a second computer.

45. A multimode information processing system as recited in claim **44** in which said template is stored in said first computer.

46. A multimode information processing system as recited in claim **44** in which said template is stored in said second computer.

47. A multimode information processing system as recited in claim **41** in which said document or file is on a first computer, and said application program is on a second computer.

48. A multimode information processing system as recited in claim **47** in which said template is stored in said first computer.

49. A multimode information processing system as recited in claim **47** in which said template is stored in said second computer.

50. A multimode information processing system as recited in claim 42 in which said document or file is on a first computer, and said application program is on a second computer.

51. A multimode information processing system as recited in claim 50 in which said template is stored in said first computer.

52. A multimode information processing system as recited in claim 50 in which said template is stored in said second computer.

53. A multimode information system as recited in claim 40, 41, or 42 in which said document or file, said at least one application program and/or said template is/are distributed among more than one computer.

54. A method of doing business comprising the steps of: inputting information from a document of file on a computer into at least one application program according to transmission format instructions;

processing the information in:

- a. a definition mode wherein content instructions, at least one of which is not a location of said information derived from a pre-scanned image of a blank form are used to define input information from within said document or file required by said at least one application program; and
- b. an extraction mode to parse at least a portion of said document or file to automatically extract at least one field of information required by said at least one application program and to transfer said at least one field of information to said at least one application program.

55. A multimode information processing system as recited in claim 54, wherein said definition mode is operative to store said content instructions in said system in association with said extraction mode.

56. multimode information processing system as recited in claim 54, wherein said definition mode is operative to store said content instructions in said system as a template for use in said extraction mode.

57. A multimode information processing system as recited in claim 54, wherein said extraction mode is operative to match at least a portion of said document or file with a template created with or available to said system.

58. A multimode information processing system as recited in claim 55, wherein said extraction mode is operative to match at least a portion of said document or file with a template created with or available to said system.

59. A multimode information processing system as recited in claim 54 in which said document or file is on a first computer, and said application program is on a second computer.

60. A multimode information processing system as recited in claim 56 in which said document or file is on a first computer, and said application program is on a second computer.

61. A multimode information processing system as recited in claim 60 in which said template is stored in said first computer.

62. A multimode information processing system as recited in claim 60 in which said template is stored in said second computer.

63. A multimode information processing system as recited in claim 57 in which said document or file is on a first computer, and said application program is on a second computer.

64. A multimode information processing system as recited in claim 63 in which said template is stored in said first computer.

65. A multimode information processing system as recited in claim 63 in which said template is stored in said second computer.

66. A multimode information processing system as recited in claim 58 in which said document or file is on a first computer, and said application program is on a second computer.

67. A multimode information processing system as recited in claim 66 in which said template is stored in said first computer.

68. A multimode information processing system as recited in claim 66 in which said template is stored in said second computer.

69. A multimode information system as recited in claim 56, 57, or 58 in which said document or file, said at least one application program and/or said template is/are distributed among more than one computer.

70. A method of processing information comprising the steps of:

inputting information from a document of file on a computer into at least one application program according to transmission format instructions;

processing the information in at least one of:

- a. a definition mode wherein content instructions, at least one of which is not a location of said information derived from a pre-scanned image of a blank form, are used to define input information from within said document or file required by said at least one application program; and
- b. an extraction mode to parse at least a portion of said document or file to automatically extract at least one field of information required by said at least one application program and to transfer said at least one field of information to said at least one application program.

71. A multimode information processing system as recited in claim 70, wherein said definition mode is operative to store said content instructions in said system in association with said extraction mode.

72. A multimode information processing system as recited in claim 70, wherein said definition mode is operative to store said content instructions in said system as a template for use in said extraction mode.

73. A multimode information processing system as recited in claim 70, wherein said extraction mode is operative to match at least a portion of said document or file with a template created with or available to said system.

74. A multimode information processing system as recited in claim 71, wherein said extraction mode is operative to match at least a portion of said document or file with a template created with or available to said system.

75. A multimode information processing system as recited in claim 70 in which said document or file is on a first computer, and said application program is on a second computer.

76. A multimode information processing system as recited in claim 72 in which said document or file is on a first computer, and said application program is on a second computer.

77. A multimode information processing system as recited in claim 76 in which said template is stored in said first computer.

78. A multimode information processing system as recited in claim 76 in which said template is stored in said second computer.

79. A multimode information processing system as recited in claim 73 in which said document or file is on a first computer, and said application program is on a second computer.

21

80. A multimode information processing system as recited in claim **79** in which said template is stored in said first computer.

81. A multimode information processing system as recited in claim **79** in which said template is stored in said second computer.

82. A multimode information processing system as recited in claim **74** in which said document or file is on a first computer, and said application program is on a second computer.

83. A multimode information processing system as recited in claim **82** in which said template is stored in said first computer.

84. A multimode information processing system as recited in claim **82** in which said template is stored in said second computer.

85. A multimode information system as recited in claim **72, 73, or 74** in which said document or file, said at least one application program and/or said template is/are distributed among more than one computer.

86. A method of processing information comprising the steps of:

inputting information from a document of file on a computer into at least one application program according to transmission format instructions;

processing the information in:

- a. a definition mode wherein content instructions, at least one of which is not a location of said information derived from a pre-scanned image of a blank form, are used to define input information from within said document or file required by said at least one application program; and
- b. an extraction mode to parse at least a portion of said document or file to extract at least one field of information required by said at least one application program and to automatically transfer said at least one field of information to said at least one application program.

87. A multimode information processing system as recited in claim **86**, wherein said definition mode is operative to store said content instructions in said system in association with said extraction mode.

88. A multimode information processing system as recited in claim **86**, wherein said definition mode is operative to store said content instructions in said system as a template for use in said extraction mode.

89. A multimode information processing system as recited in claim **86**, wherein said extraction mode is operative to

22

match at least a portion of said document or file with a template created with or available to said system.

90. A multimode information processing system as recited in claim **87** wherein said extraction mode is operative to match at least a portion of said document or file with a template created with or available to said system.

91. A multimode information processing system as recited in claim **86** in which said document or file is on a first computer, and said application program is on a second computer.

92. A multimode information processing system as recited in claim **88** in which said document or file is on a first computer, and said application program is on a second computer.

93. A multimode information processing system as recited in claim **92** in which said template is stored in said first computer.

94. A multimode information processing system as recited in claim **76** in which said template is stored in said second computer.

95. A multimode information processing system as recited in claim **89** in which said document or file is on a first computer, and said application program is on a second computer.

96. A multimode information processing system as recited in claim **95** in which said template is stored in said first computer.

97. A multimode information processing system as recited in claim **95** in which said template is stored in said second computer.

98. A multimode information processing system as recited in claim **90** in which said document or file is on a first computer, and said application program is on a second computer.

99. A multimode information processing system as recited in claim **98** in which said template is stored in said first computer.

100. A multimode information processing system as recited in claim **98** in which said template is stored in said second computer.

101. A multimode information system as recited in claim **88, 89, or 90** in which said document or file, said at least one application program and/or said template is/are distributed among more than one computer.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,683,697 B1
DATED : January 27, 2004
INVENTOR(S) : Robert Lech, Mitchell A. Medina and Catherine B. Elias

Page 1 of 2

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 15,

Line 48, after "to" insert -- customizable --.
Line 50, after "instructions" delete -- , at least one of which is not a location of said information derived from a pre-scanned image of a blank form, --.
Line 61, after, "program" insert -- according to said customizable transmission format instructions --

Column 16,

Line 51, after "to" insert -- customizable --.

Column 15,

Line 53, after "instructions" delete -- , at least one of which is not a location of said information derived from a pre-scanned image of a blank form, --.
Line 64, after "program" insert -- according to said customizable transmission format instructions --.

Column 17,

Line 54, after "instructions" delete -- , at least one of which is not a location of said information derived from a pre-scanned image of a blank form, --.
Line 61, after "on" insert -- customizable --.
Line 65, after "mode" insert -- according to said customizable transmission format instructions --.

Column 18,

Line 14, after "to" insert -- customizable --.
Line 16, after "instructions" delete -- , at least one of which is not a location of said information derived from a pre-scanned image of a blank form --.
Line 27, after "program" insert -- according to said customizable transmission format instructions --.

Column 19,

Line 18, after "to" insert -- customizable --.
Line 20, after "instructions" delete -- , at least one of which is not a location of said information derived from a pre-scanned image of a blank form, --.
Line 31, after "program" insert -- according to said customizable transmission format instructions --.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,683,697 B1
DATED : January 27, 2004
INVENTOR(S) : Robert Lech, Mitchell A. Medina and Catherine B. Elias

Page 2 of 2

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 20,

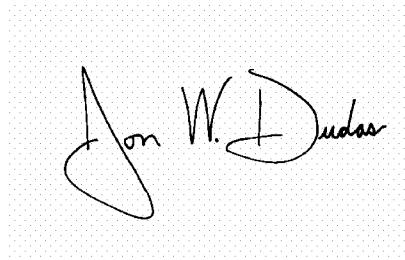
Line 22, after "to" insert -- customizable --.
Line 24, after "instructions" delete -- , at least one of which is not a location of said information derived from a pre-scanned image of a blank form --.
Line 34, after "program" insert -- according to said customizable transmission format instructions. --

Column 21,

Line 25, after "to" insert -- customizable --.
Line 27, after "instructions" delete -- , at least one of which is not a location of said information derived from a pre-scanned image of a blank form, --.
Line 38, after "program" insert -- according to said customizable transmission format instructions. --

Signed and Sealed this

Twenty-sixth Day of July, 2005

A handwritten signature in black ink, appearing to read "Jon W. Dudas", is placed within a dotted rectangular box.

JON W. DUDAS
Director of the United States Patent and Trademark Office



US007075673B2

(12) **United States Patent**
Lech et al.

(10) **Patent No.:** **US 7,075,673 B2**
(b5) **Date of Patent:** **Jul. 11, 2006**

(54) **INFORMATION PROCESSING
METHODOLOGY**

(75) Inventors: **Robert Lech**, Jackson, NJ (US);
Mitchell A. Medina, New York, NY (US); **Catherine B. Elias**, Plainsboro, NJ (US)

(73) Assignee: **EON-Net L.P.**, Tortola (VG)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 34 days.

(21) Appl. No.: **10/704,484**

(22) Filed: **Nov. 6, 2003**

(65) **Prior Publication Data**

US 2004/0070793 A1 Apr. 15, 2004

Related U.S. Application Data

(60) Continuation of application No. 09/458,162, filed on Dec. 9, 1999, now Pat. No. 6,683,697, which is a continuation of application No. 09/044,159, filed on Mar. 19, 1998, now Pat. No. 6,094,505, which is a continuation of application No. 08/487,150, filed on Jun. 7, 1995, now Pat. No. 5,768,416, which is a division of application No. 08/348,224, filed on Nov. 28, 1994, now Pat. No. 5,625,465, which is a continuation of application No. 08/143,135, filed on Oct. 29, 1993, now Pat. No. 5,369,508, which is a continuation of application No. 07/672,865, filed on Mar. 20, 1991, now Pat. No. 5,258,855.

(51) **Int. Cl.**
H04N 1/40 (2006.01)

(52) **U.S. Cl.** **358/1.15; 382/175**
(58) **Field of Classification Search** **358/1.15; 382/175, 177, 180, 282, 287, 306, 317**

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,013,718 A	12/1961	Shepard et al.
3,200,372 A	8/1965	Hamburgen
3,303,463 A	2/1967	Hamburgen
3,434,110 A	3/1969	Bucklin, Jr. et al.
3,492,653 A	1/1970	Fosdick et al.
3,582,883 A	6/1971	Shepard
3,582,884 A	6/1971	Shepard

(Continued)

FOREIGN PATENT DOCUMENTS

EP 0 107 083 B1 7/1988

(Continued)

OTHER PUBLICATIONS

PC Magazine, vol. 5, No. 16, Sep. 30, 1986.

(Continued)

Primary Examiner—Thomas D. Lee

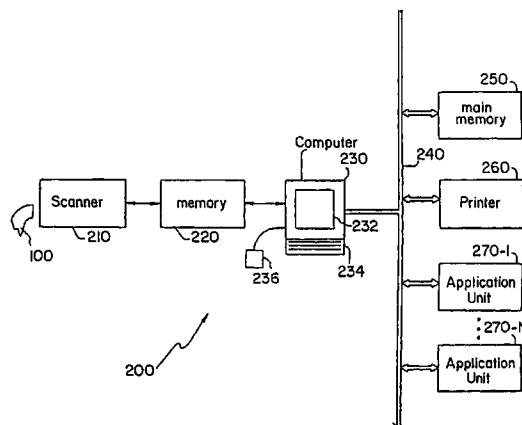
Assistant Examiner—Stephen Brinich

(74) Attorney, Agent, or Firm—Foley & Lardner LLP

(57) **ABSTRACT**

An information processing methodology gives rise to an application program interface which includes an automated digitizing unit, such as a scanner, which inputs information from a diversity of hard copy documents and stores information from the hard copy documents into a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which designate portions of the stored document information required by a particular application program. The selected stored document information is then placed into the transmission format required by a particular application program in accordance with transmission format instructions. After the information has been transmission formatted, the information is transmitted to the application program. In one operational mode, the interface interactively prompts the user to identify, on a display, portions of the hard copy documents containing information used in application programs or for storage.

57 Claims, 15 Drawing Sheets



US 7,075,673 B2

Page 2

U.S. PATENT DOCUMENTS

3,584,144 A	6/1971	Shepard et al.
3,631,396 A	12/1971	Spertus
3,832,682 A	8/1974	Brok et al.
3,848,228 A	11/1974	MacNeill
3,903,517 A	9/1975	Hafner
RE29,104 E	1/1977	Shepard
4,021,777 A	5/1977	Shepard
4,034,343 A	7/1977	Wilmer
4,041,454 A	8/1977	Shepard et al.
4,047,154 A	9/1977	Vitols et al.
4,132,978 A	1/1979	Mercier
4,387,964 A	6/1983	Arrazola et al.
4,553,261 A	11/1985	Froessl
4,564,752 A	1/1986	Lepic et al.
4,572,962 A	2/1986	Shepard
4,593,367 A	6/1986	Slack et al.
4,659,940 A	4/1987	Shepard
4,667,248 A	5/1987	Kanno
4,672,678 A	6/1987	Koezuka et al.
4,760,246 A	7/1988	Shepard
4,760,606 A	7/1988	Lesnick et al.
4,776,016 A *	10/1988	Hansen 704/275
4,782,509 A	11/1988	Shepard
4,802,104 A	1/1989	Ogiso
4,802,231 A	1/1989	Davis
4,803,734 A	2/1989	Onishi et al.
4,931,957 A	6/1990	Takagi et al.
4,933,979 A	6/1990	Suzuki et al.
4,974,260 A	11/1990	Rudak
5,017,763 A	5/1991	Shepard
5,031,121 A	7/1991	Iwai et al.
5,034,990 A	7/1991	Klees
5,052,038 A	9/1991	Shepard
5,095,445 A	3/1992	Sekiguchi
5,140,139 A	8/1992	Shepard
5,140,650 A	8/1992	Casey et al.
5,153,927 A	10/1992	Yamanari
5,159,667 A	10/1992	Borrey et al.
5,191,525 A	3/1993	LeBrun et al.
5,218,539 A	6/1993	Elphick et al.
5,228,100 A	7/1993	Takeda et al.
5,245,166 A	9/1993	Shepard
5,251,268 A	10/1993	Colley et al.
5,257,328 A	10/1993	Shimizu
5,258,855 A	11/1993	Lech et al.
5,282,267 A	1/1994	Woo, Jr. et al.
5,307,424 A	4/1994	Kuehl
5,367,619 A	11/1994	Dipaolo
5,404,294 A	4/1995	Kamik
5,416,849 A	5/1995	Huang
5,444,840 A	8/1995	Froessl
5,448,738 A	9/1995	Good et al.
5,452,379 A	9/1995	Poor
5,455,875 A	10/1995	Chevion et al.
5,506,697 A	4/1996	Li et al.
5,511,135 A	4/1996	Rhyne et al.
5,526,447 A	6/1996	Shepard
5,550,930 A	8/1996	Berman et al.
5,555,325 A	9/1996	Burger
5,696,854 A	12/1997	Shepard
5,734,761 A	3/1998	Bagley
5,852,685 A	12/1998	Shepard
5,923,792 A	7/1999	Shyu et al.
5,933,531 A	8/1999	Lorie
6,067,517 A *	5/2000	Bahl et al. 704/256
6,094,505 A	7/2000	Lech et al.

FOREIGN PATENT DOCUMENTS

JP	64-38883	2/1989
JP	3-161886	7/1991

OTHER PUBLICATIONS

TopScan Professional User's Guide, Calera Recognition Systems, pp. v-vii, xi-xiii, and 1-6, 1989.

Que's Computer user's Dictionary, 2nd Ed., Bryan Pfaffenberger (author); 1991; p. 144.

"Kurzwell 5200 Intelligent Scanning System", Xerox Imaging Systems, Inc., 1990.

Edward O. Welles, *Decisions, Decisions*, Inc., Aug. 1990, pp. 80-90.

Palantir PagePro User's Guide, Rev. A, Dec. 1986, sections 1 and 4 and Appendix C.

PageRead Library Developer's Guide, Rev. B, Aug. 1989, pp. PRL1-3, SHC25-26, TUT9-10.

With AutoClass there is no more to Indexing than OCR, Remittance and Document Processing Today, Nov.-Dec. 1990, p. 46.

Daniel Borrey, *Machine Recognition and Classification of Documents*, Remittance and Document Processing Today, Nov.-Dec. 1990, pp. 20-23.

The Very Best in Optical Character Recognition, IMAGING, Mar. 1992, pp. 43-47.

Why Insurance Companies Take the Risk on Document Imaging, IMAGING, Mar. 1992, pp. 48-54.

"OCR for Forms" (advertisement), IMAGING, Apr. 1992.

David Black, *The Right and The Wrong Ways to Index*, IMAGING, May 1992, pp. 47-50.

Greg Bartels, *How to Successfully Convert Your Backfiles*, IMAGING, May 1992, pp. 55-56.

But, Is It a Boy or Girl? IMAGING, Oct. 1992, p. 10.

Gerry Frieser, *Suddenly, OCR is a "Must Buy"*, IMAGING, Dec. 1992, pp. 22-25.

14 of the Hottest OCR Software Packages and Systems, IMAGING, Dec. 1992, pp. 26-31, 33, 35.

How Form Processing Works, Plus Pros and Cons, IMAGING, Dec. 1992, p. 36.

Forms Processing Products Meet the Challenge of OCRing Forms, IMAGING, Dec. 1992, p. 38-40.

Herbert F. Schantz, *Forms Automation and Integrated Imaging (OCR) Systems*, Remittance and Document Processing Today, Mar. -Apr. 1991, pp. 9-11.

Don Merz, *OCR: A Health Insurance Application*, Remittance and Document Processing Today, Jul.-Aug. 1989, pp. 18-20.

Industry News, Remittance and Document Processing Today, Jul.-Aug. 1989, p. 22.

Industry News and New Products, Remittance and Document Processing Today, Oct. 1984, pp. 17-18.

R. C. Gonzalez, *Designing Balance Into An OCR System*, Remittance and Document Processing Today, Mar. 1988, pp. 7, 10-11.

Ambrose R. Rightler, *OCR Qualtiy Control Procedures for Remittance Processing: Can You Afford to be Without It?* Remittance and Document Processing Today, Mar. 1988, pp. 12-15.

Product Watch, MACWEEK, Oct. 3, 1989, pp. 32, 34, 38, 40, 42.

Industry News, Remittance and Document Processing Today, Jan. 1989, p. 9.

Scott Beamer, *Mac OCR Takes a Big Step Forward*, MACWEEK, Jun. 13, 1989.

Matthew Lake, *Strength of Character (Recognition)*, PUBLISH, Jan. 1991, pp. 62-67.

R. David Nelson and Karen A. Hamill, *Optical Scanning at Chemical Abstracts Service for Building Computer Files*

US 7,075,673 B2

Page 3

From Printed Index Data, Recognition Technologies Today, Feb. 1985, pp. 1-6, 15.

Gerald Farmer, *HNC IDEPT9 and Recognition Enhanced Data Entry: The Cost-Cutting Approach to Automated Data Entry*, Remittance and Document Processing Today, Jan.-Feb. 1991, pp. 24-26.

David Gertler, *Automated Data Entry*, Seybold Report on Desktop Publishing, Jan. 15, 1990, pp. 3-17.

Eric Aas and Peter Davidoff, *Teaching Your Scanner to Read*, Personal Publishing, May 1990, pp. 28, 31, 33.

Phillip Robinson et al., *Character Witnesses*, MACUSER, Jul. 1990, pp. 120-136.

Brita Meng, *Text Without Typing*, MACWORLD, Oct. 1990, pp. 177-183.

Jim Heid, *Getting Started with Optical Character Recognition*, MACWORKD, Oct. 1990, pp. 297-301.

Stanford Diehl and Howard Eglowstein, *Tame the Paper Tiger*, BYTE, Apr. 1991, pp. 220-238.

Alan Joch and Rich Graham, *Voices of Experience*, BYTE, Apr. 1991, pp. 239-241.

Gregory Boleslavsky and Roman Tutunikov, *The New Generation of OCR*, INFORM, Jan. 1990, pp. 34-37.

Calera Recognition Systems, Inc., TopScan Professional User's Guide: Complete Document Recognition for PCs and Compatibles (1989).

Calera Recognition Systems, Inc., TopScan Professional Installation Notes for Scanners, Fax Cards, and System Configuration (1989).

Calera Recognition Systems, Inc., TopScan Professional Troubleshooting Guide (1989).

Xerox Imaging Systems, Inc., Datacopy AccuText User's Guide (1989).

Invoice from Corporate Intelligence Corporation to Workman, Nydegger & Seeley, dated Sep. 21, 1999.

Examiner Interview Summary Record, Jan. 25, 1996, and related Amendment, for U.S. Appl. No. 08/097,131.

Jane B. Newman, Formstar Ad; "Stack the Facts, Not the Forms"—May 24, 1987; "Stack and Send Just the Facts—and Improve the Efficiency of your Forms Application".

TeleImage Systems Document and Image Database Systems User's Manual- Ramat Gan, Israel; Table of Contents and pp. 2-1 through 5-20.

Form Out! Programmer's Manual; TeleImage Systems, Ramat Gan Israel; published Feb. 1991 pp. I, II, III, IV, V, VI, VII; 1.1-6.29 and A.1-G.11.

IPLEXUS, "The Extended Data Processing Concept", Plexus Computers, Ltd., 1988.

Laserfiche United, Capabilities and Functional Specifications, 2002 Compulink Management Center, Inc.

The LaserFiche™ System, LF00007-LF00010, Copyright 1998 Compulink Management Center.

LaserFiche™, Integrated Document Imaging and Text Processing Systems, Compulink Management Center, Inc.

Compulink Management Center, Copyright 1989, LF-00006.

PC Week, Document-Archival System Combines DOS with CD ROM, LF00011-LF-00032.

LASERFICHE, Zone OR Plug-in, 2002 Compulink Management Center, Inc.

LASERFICHE, Efficient Integration, <http://www.laserfiche.com/products/techadmin/integration.html>, Jul. 2, 2004.

LASERFICHE, Integrator's Toolkit, Compulink Management Center, Inc. 2003.

LASERFICHE, LaserFiche Integration Express-GIS™, <http://www.laserfiche.com/products/integrationexpress-gis.html>, Jul. 2, 2004.

OCR Texiris, Texiris User's Guide, Version 2.1, Image Recognition Integrated systems S.A.

Texiris User's Guide, 13. Interfacing Data Base Management Systems.

Texiris User's Guide, 5. Correction Tools.

* cited by examiner

U.S. Patent

Jul. 11, 2006

Sheet 1 of 15

US 7,075,673 B2

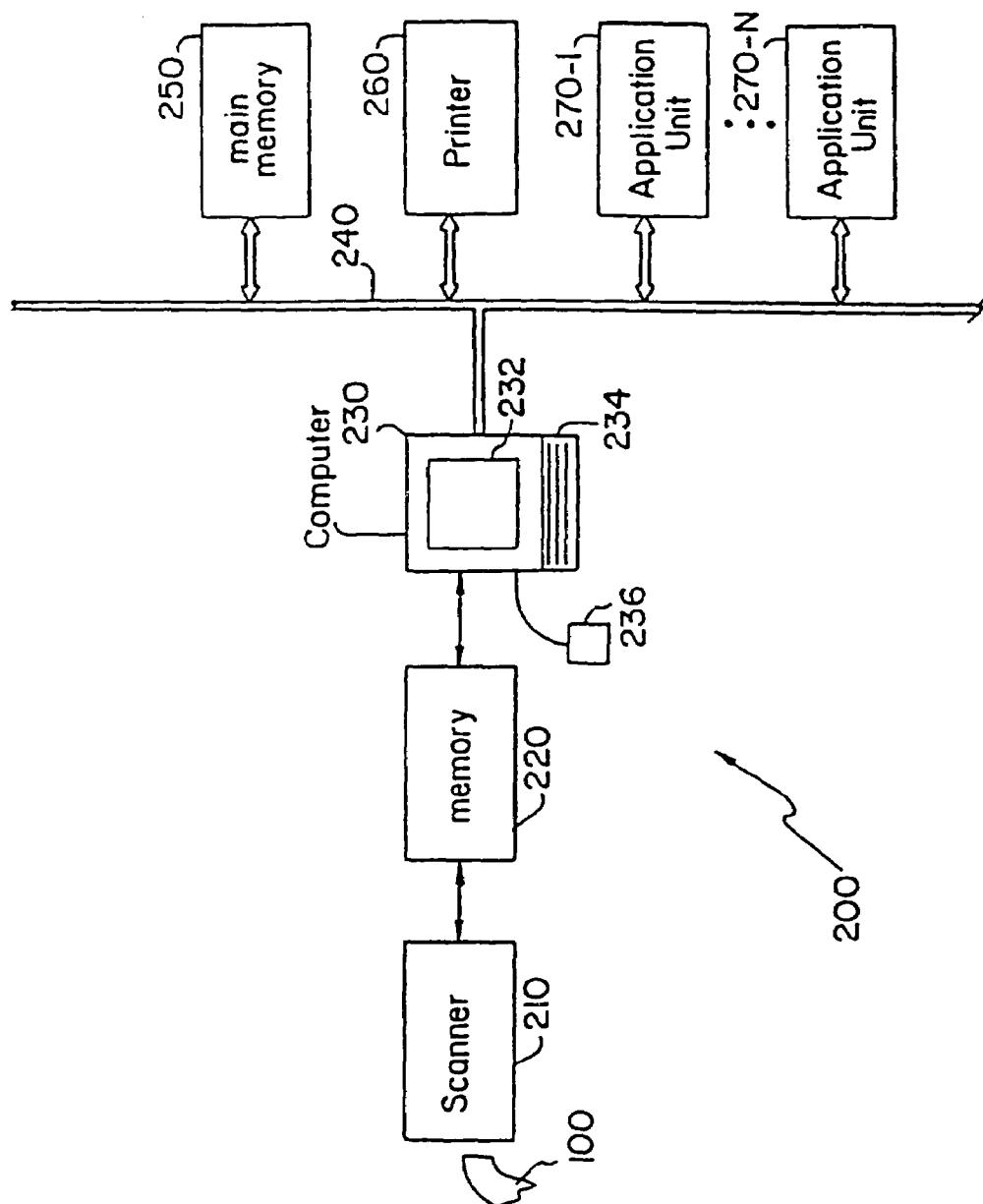


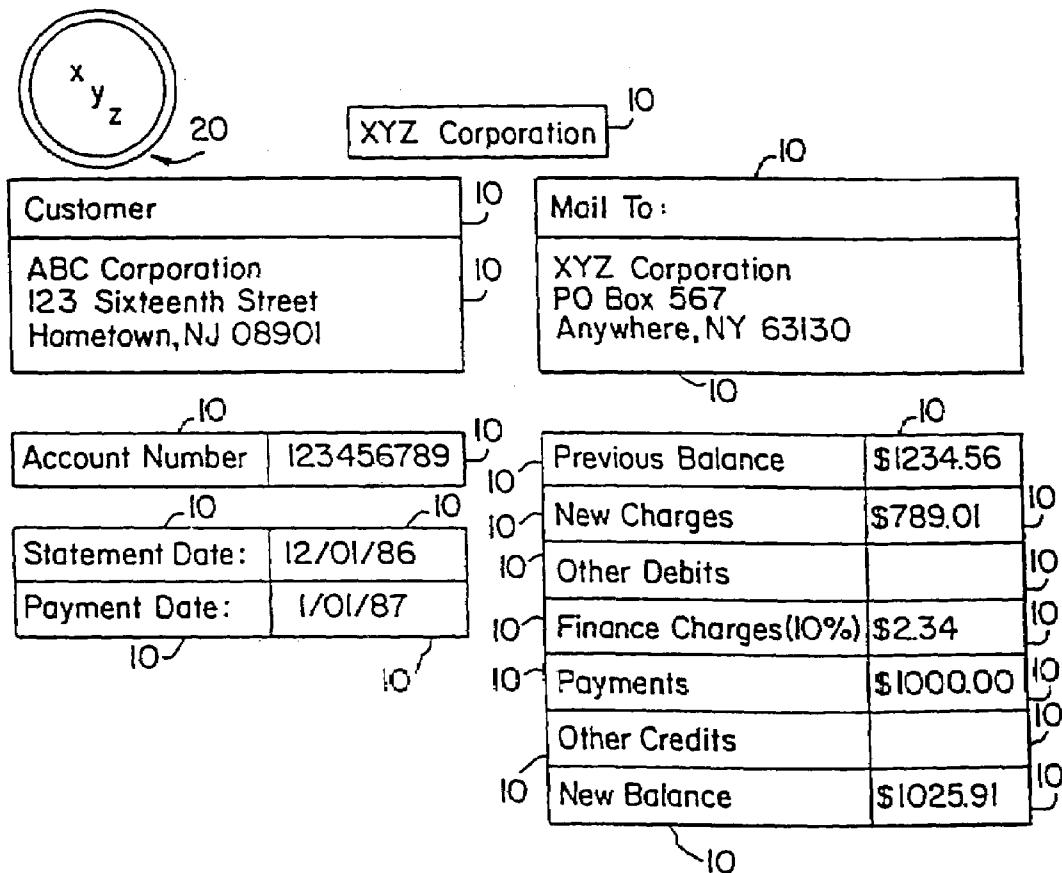
FIG. 1

U.S. Patent

Jul. 11, 2006

Sheet 2 of 15

US 7,075,673 B2



Winter Sale In Effect Throughout January!

30

FIG. 2

U.S. Patent

Jul. 11, 2006

Sheet 3 of 15

US 7,075,673 B2

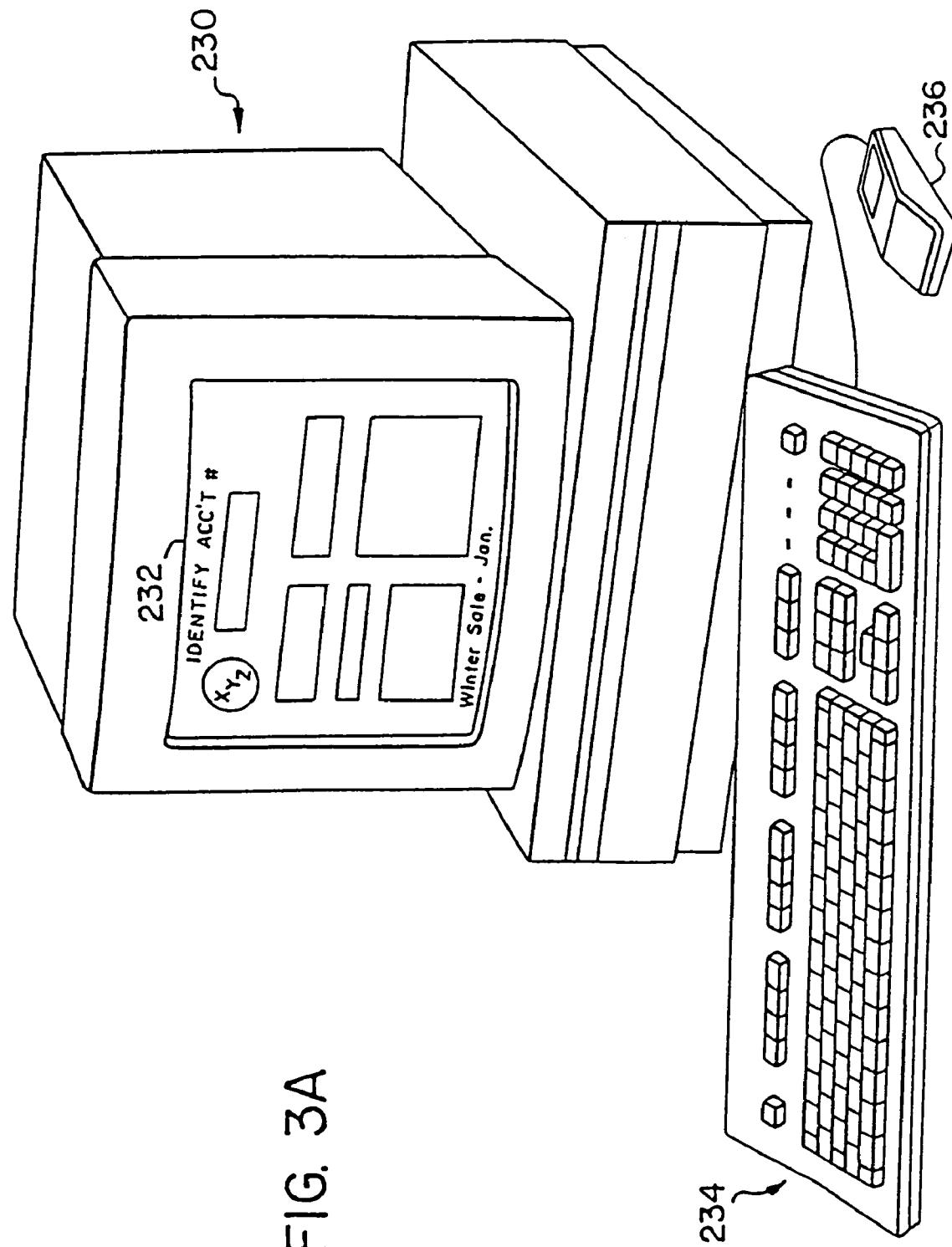


FIG. 3A

U.S. Patent

Jul. 11, 2006

Sheet 4 of 15

US 7,075,673 B2

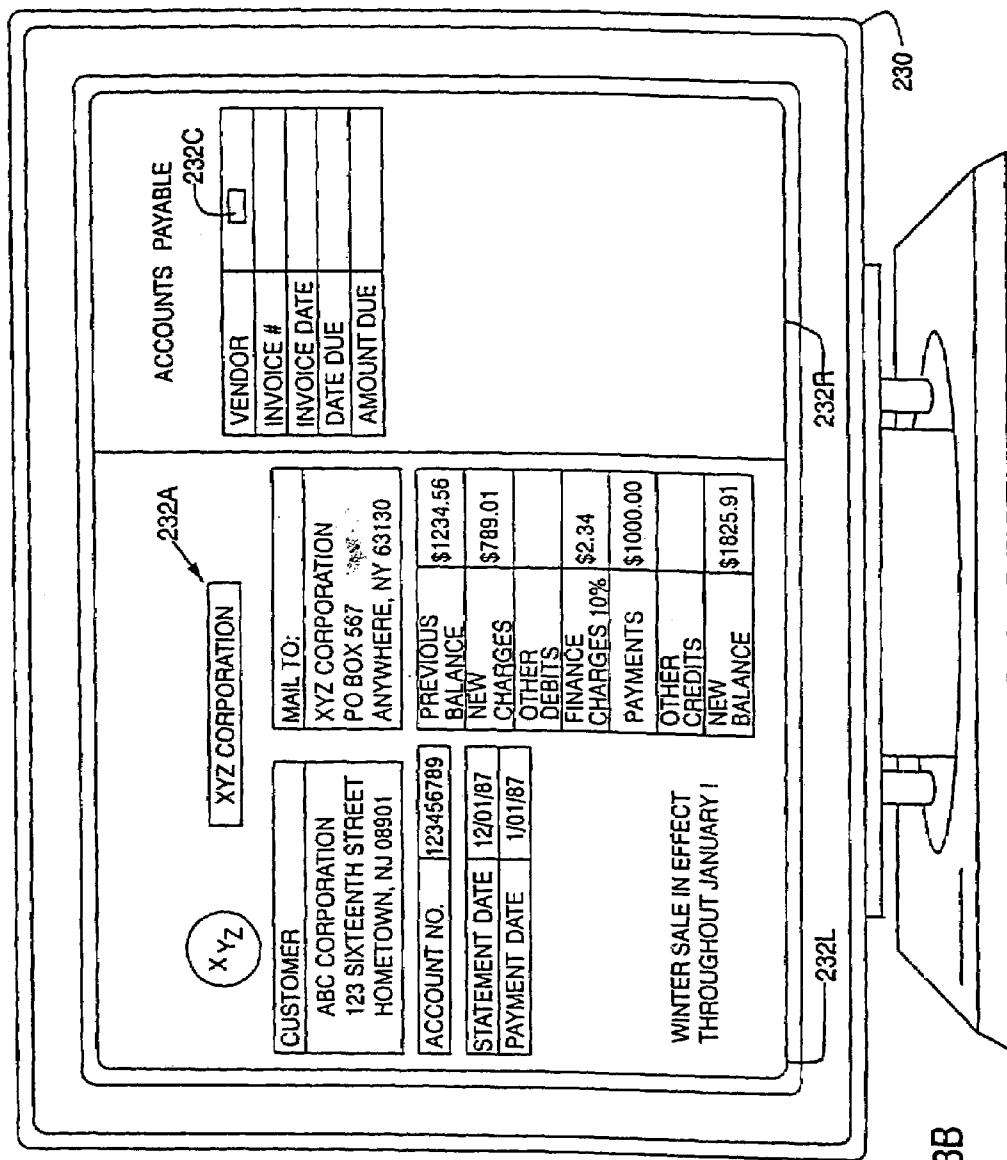


FIG. 3B

U.S. Patent

Jul. 11, 2006

Sheet 5 of 15

US 7,075,673 B2

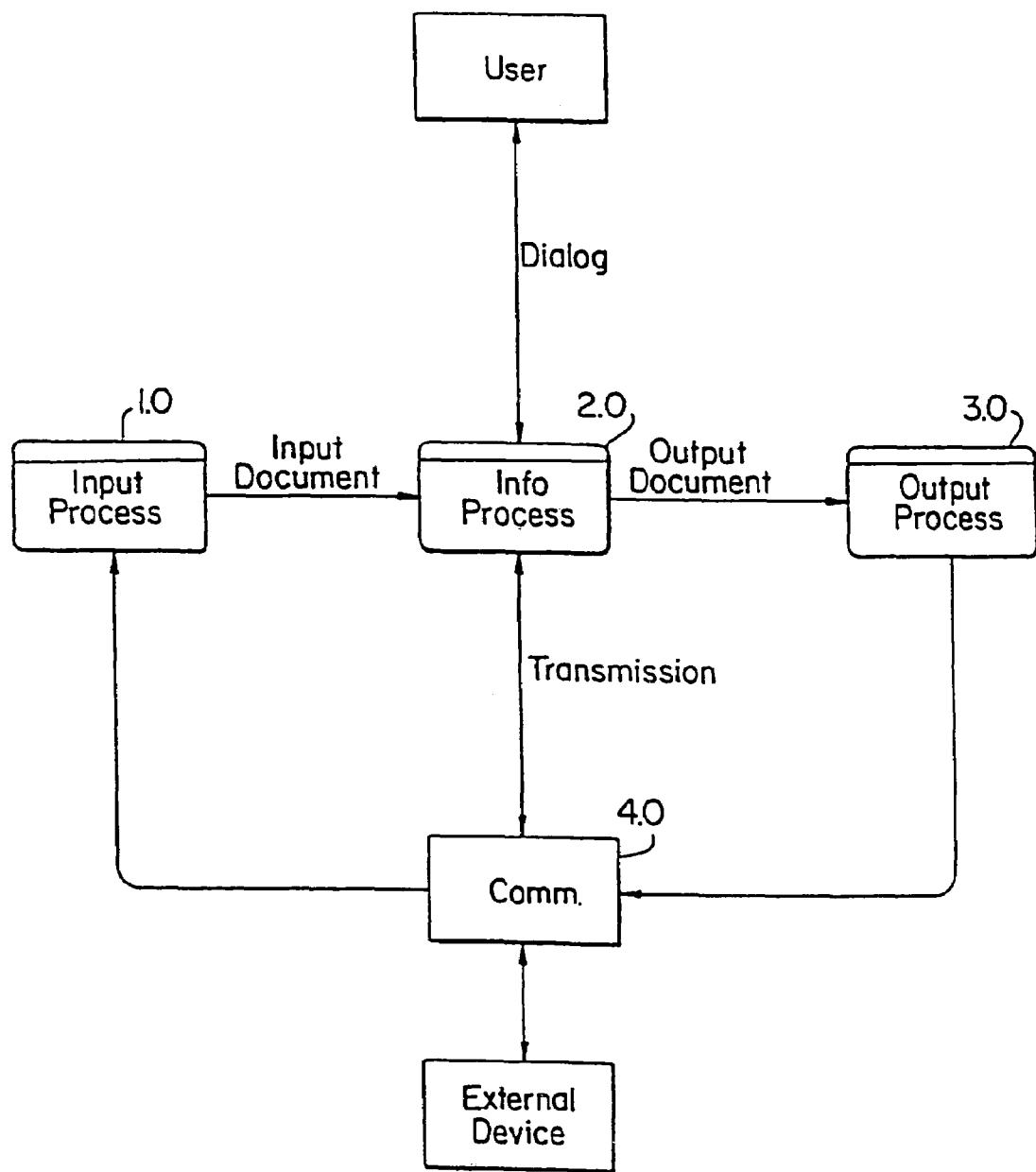


FIG. 4

U.S. Patent

Jul. 11, 2006

Sheet 6 of 15

US 7,075,673 B2

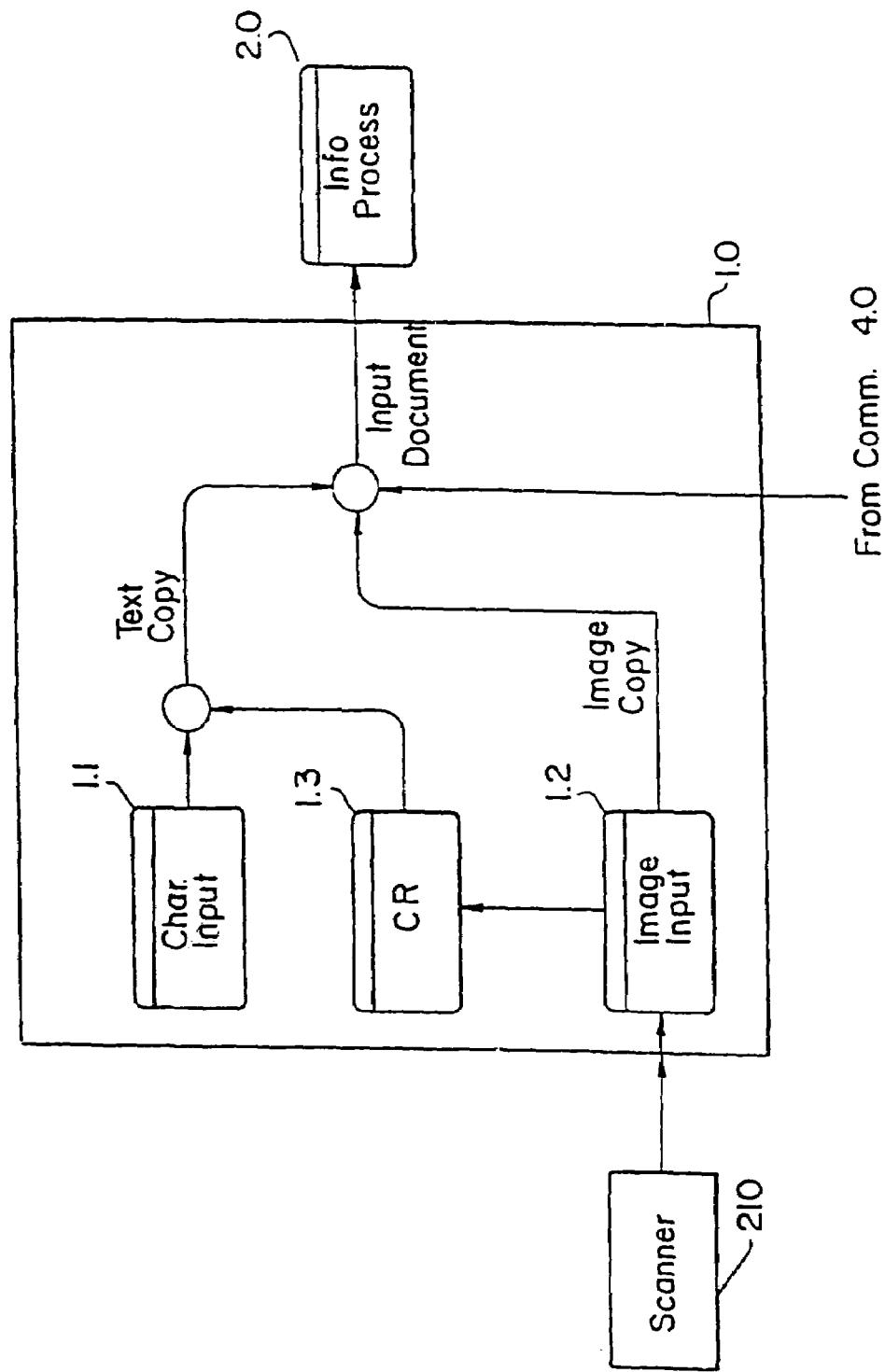


FIG. 5

U.S. Patent

Jul. 11, 2006

Sheet 7 of 15

US 7,075,673 B2

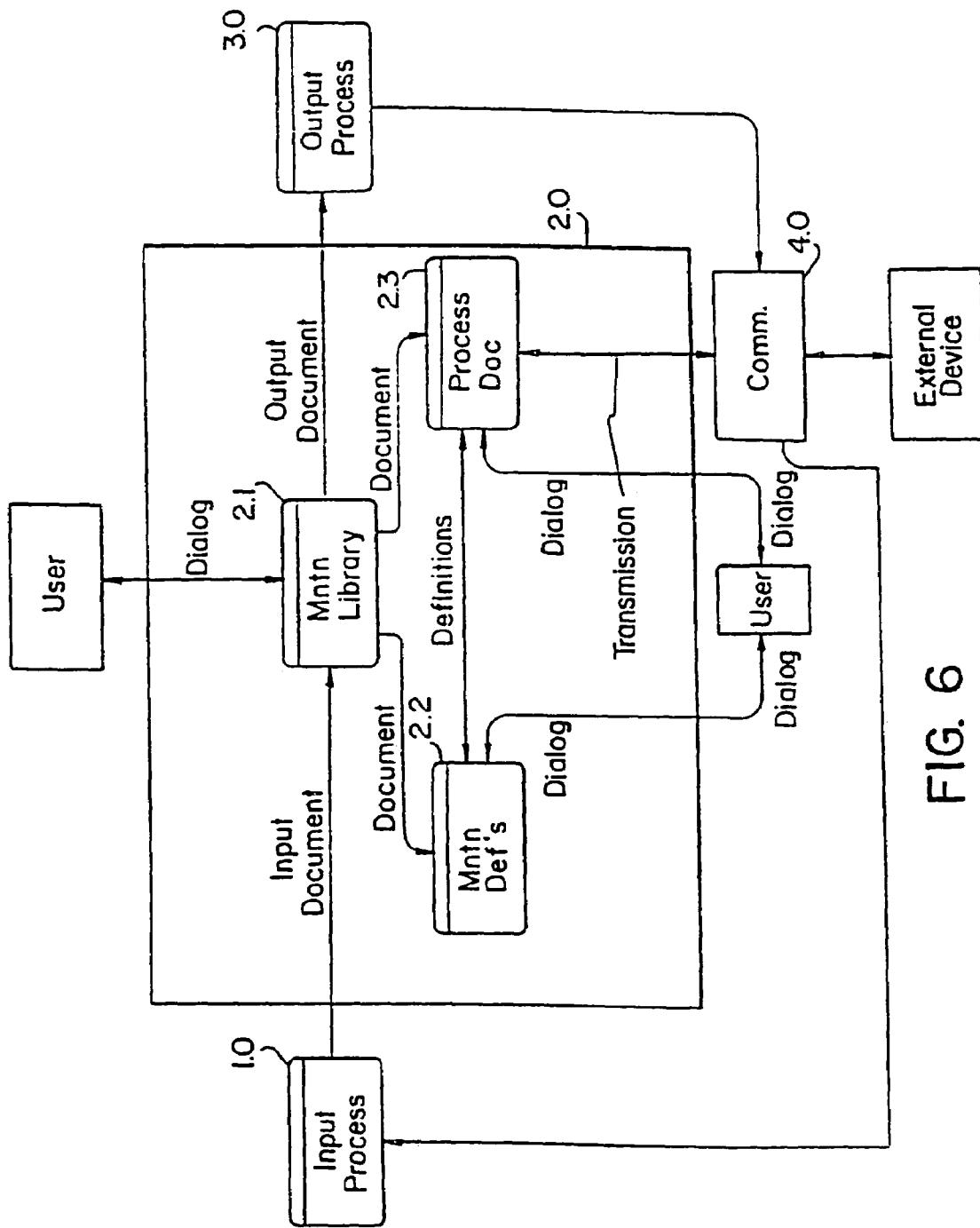


FIG. 6

U.S. Patent

Jul. 11, 2006

Sheet 8 of 15

US 7,075,673 B2

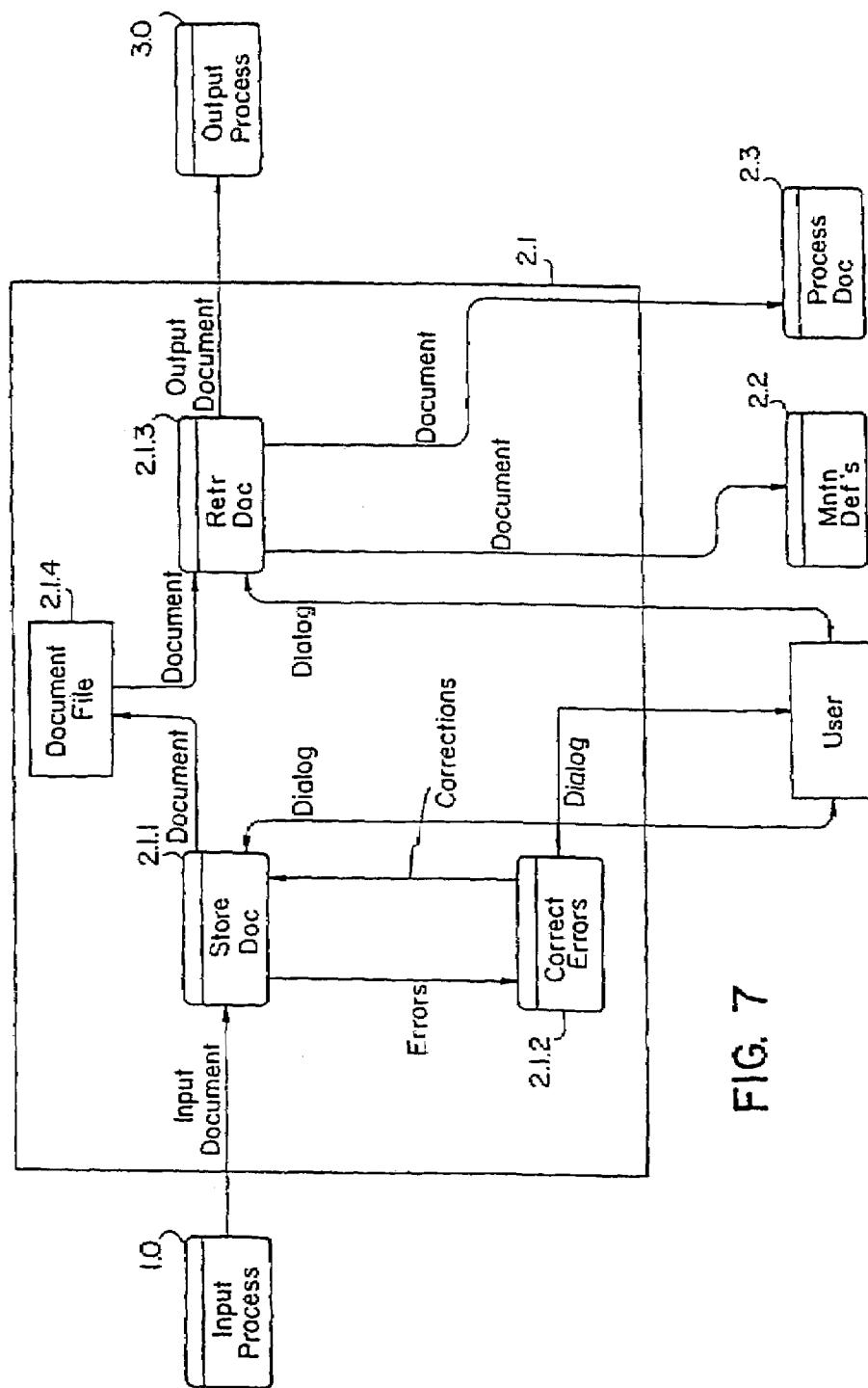


FIG. 7

U.S. Patent

Jul. 11, 2006

Sheet 9 of 15

US 7,075,673 B2

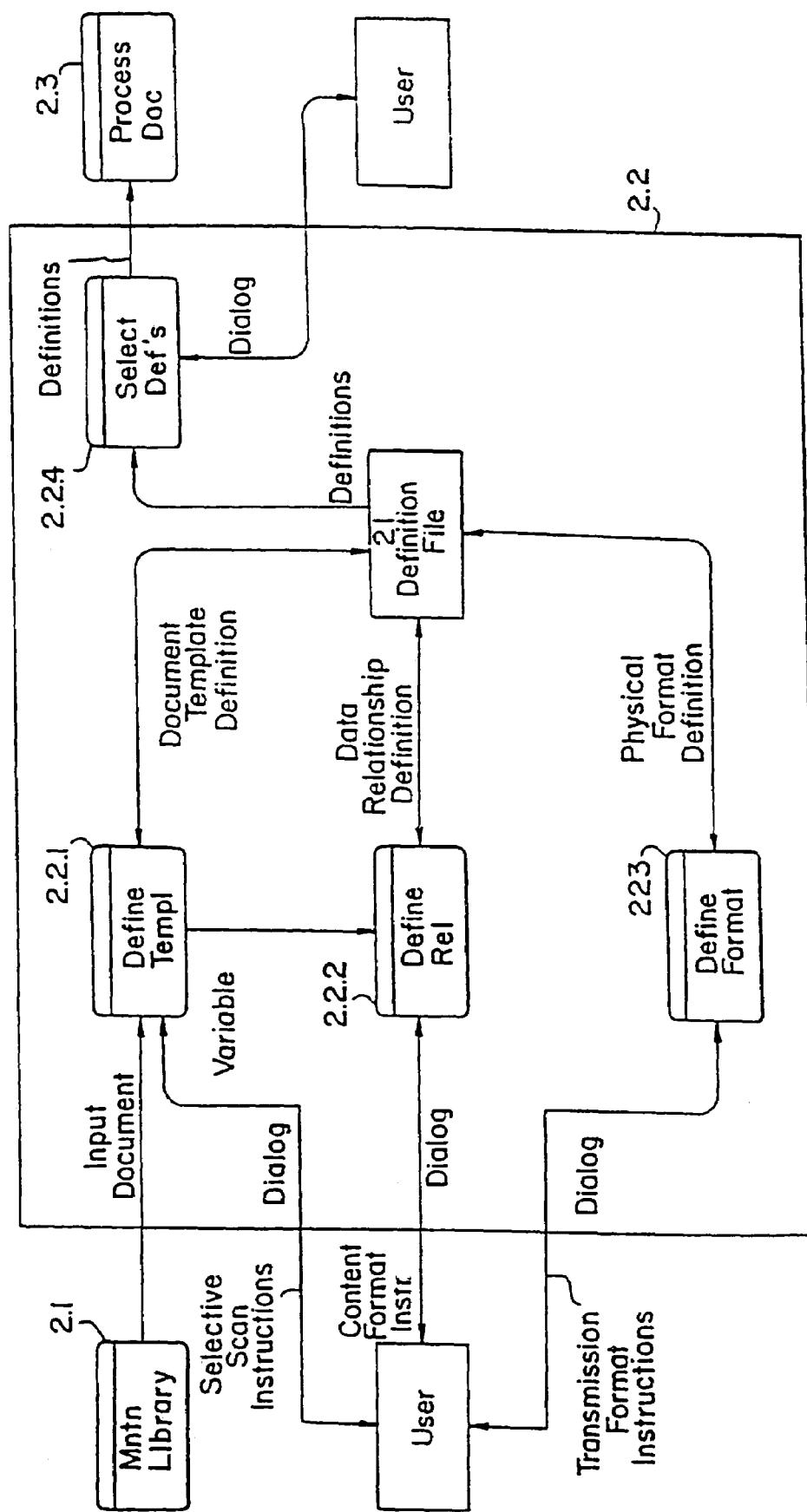


FIG. 8

U.S. Patent

Jul. 11, 2006

Sheet 10 of 15

US 7,075,673 B2

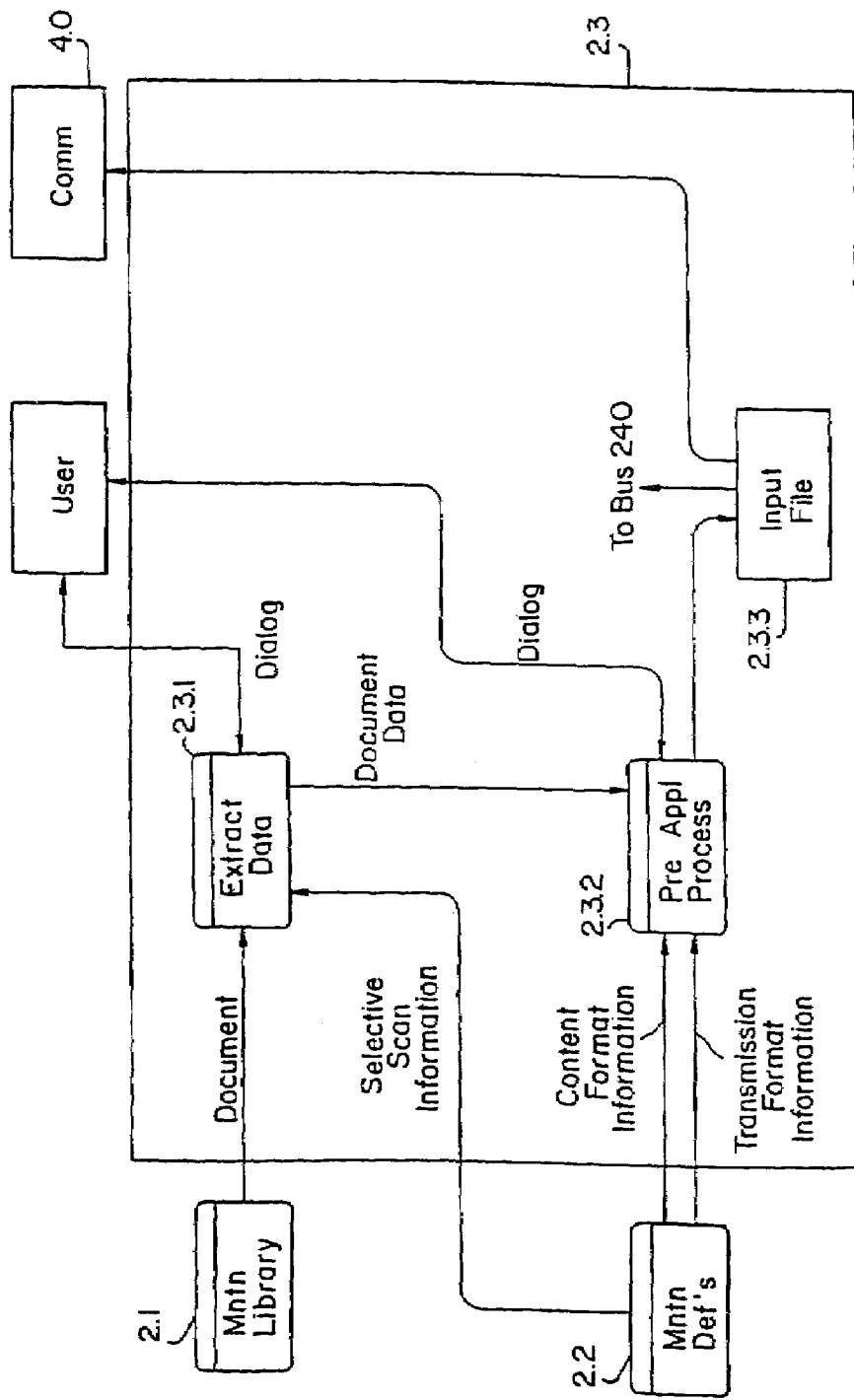


FIG. 9

U.S. Patent

Jul. 11, 2006

Sheet 11 of 15

US 7,075,673 B2

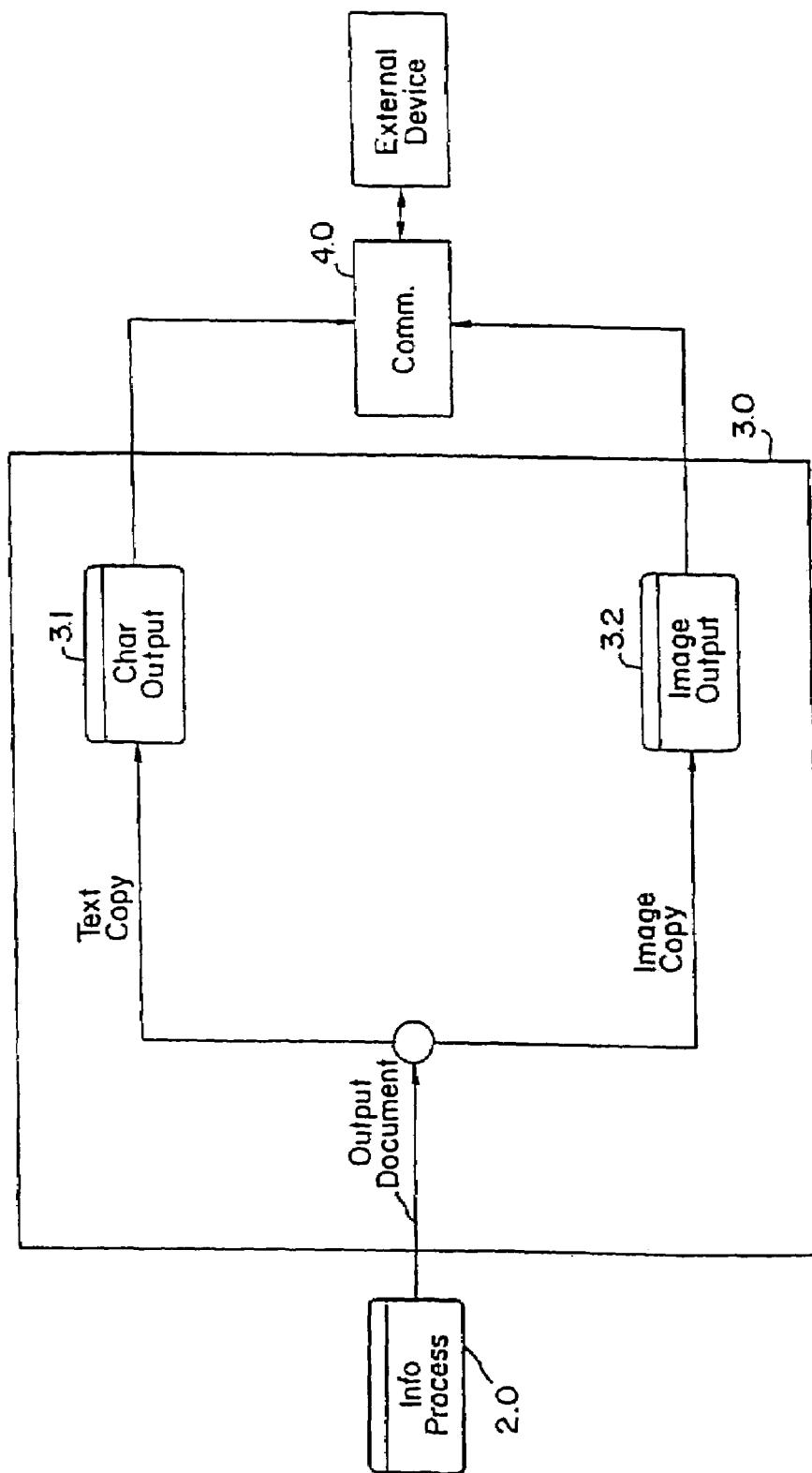


FIG. 10

U.S. Patent

Jul. 11, 2006

Sheet 12 of 15

US 7,075,673 B2

FIG. 11

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Heading 2	2
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
Other Credits	
New Balance	\$1025.91

U.S. Patent

Jul. 11, 2006

Sheet 13 of 15

US 7,075,673 B2

FIG. 12A

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
New Balance	\$1025.91

FIG. 12B

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

FIG. 12C

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

U.S. Patent

Jul. 11, 2006

Sheet 14 of 15

US 7,075,673 B2

FIG. 13A

{ > 1 > 1 > 1 > "XYZ Corporation"
> 2 > 2 > 25 > + 123456789 >
> 3 > 2 > 1 > D12 / 01 / 86 >
> 4 > 2 > 11 > D12 / 15 / 86 >
> 5 > 2 > 21 > D01 / 01 / 87 >
> 6 > 10 > 25 > \$1234.56 >
> 7 > 11 > 25 > \$789.01 >
> 8 > 13 > 25 > \$2.34 >
> 9 > 14 > 25 > \$1000.00 >
> 10 > 16 > 25 > \$1025.91 >

FIG. 13B

{ > 1 > 1 > 1 > \$1234.56 >
> 2 > 2 > 1 > "XYZ Corporation"
> 3 > 3 > 1 > "PO Box 567"
> 4 > 4 > 1 > "Anywhere, NY 63130"

FIG. 13C

{ / 1 / 1 / \$1234.56 //
/ 2 / 2 / 1 / *XYZ Corporation*
/ 3 / 3 / 1 / *PO Box 567*
/ 4 / 4 / 1 / *Anywhere, NY 63130*

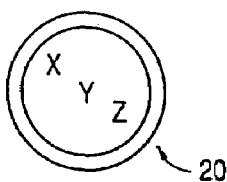
U.S. Patent

Jul. 11, 2006

Sheet 15 of 15

US 7,075,673 B2

FIG. 14



XYZ Corporation

Customer

ABC Corporation
123 Sixteenth Street
Hometown, NJ 88981

Mail To:

△ XYZ Corporation
PO Box 567
Anywhere, NY 63130 △

Account Number 123456789

Previous Balance \$1234.56

Statement Date: 12/01/86 New Charges \$789.01

Payment Date: 1/01/87

Other Debits

Finance Charges (10%) \$2.34

Payments \$1000.00

WINTER SALE IN EFFECT THROUGHOUT JANUARY:

Other Credits

New Balance \$1025.91

30

US 7,075,673 B2

1

INFORMATION PROCESSING METHODOLOGY

This application is a continuation of application Ser. No. 09/458,162, filed Dec. 9, 1999 now U.S. Pat. No. 6,683,697 which is a continuation of application Ser. No. 09/044,159, filed Mar. 19, 1998 (now U.S. Pat. No. 6,094,505), which is a continuation of application Ser. No. 08/487,150, filed Jun. 7, 1995 (now U.S. Pat. No. 5,768,416), which is a divisional of Ser. No. 08/348,224, filed Nov. 28, 1994 (now U.S. Pat. No. 5,625,465), which is a continuation of Ser. No. 08/143,135, filed Oct. 29, 1993 (now U.S. Pat. No. 5,369,508), which is a continuation of Ser. No. 07/672,865, filed Mar. 20, 1991 (now U.S. Pat. No. 5,258,855).

BACKGROUND OF THE INVENTION

The invention is directed to a system for efficiently processing information originating from hard copy documents. More specifically, the invention is directed to a hard copy document to application program interface which minimizes the need to manually process hard copy documents.

In the past, information contained on hard copy documents was manually entered into a computer via the input controller of a particular computer. The original document was then filed away for future reference. Automatic input of data was limited to the input of Magnetic Ink Character Recognition (MICR) data and to Optical Character Recognition (OCR) data. This fixed-position data was forwarded directly to a dedicated computer application specifically designed to accommodate the input format. In more recent years typewritten text has been mechanically inputted into a computer via a text file. Examples of this latter type of system are word processors and photo-typesetters.

These conventional systems have limitations which decrease the efficiency of processing information from a hard copy document. For example, the systems discussed above are limited in their application to MICR, OCR, or typewritten data. Parsing and processing data is limited to the particular requirements of the particular computer application which requires the input data. In addition, in these conventional systems, the actual hard copy document must be retained for future reference at great expense.

In a sophisticated computer network, different users may require different portions of the information contained on a hard copy document. For example, if the hard copy document is an invoice returned with payment of a bill, the accounting department may need all of the monetary information contained on the bill while the mailroom may need only customer address information, to update a customer's address. Therefore, there is a need for a system in which specific information from a hard copy document can be selectively distributed to various users.

Another problem with conventional systems is that users, even within the same company, may require that the information extracted from a hard copy document be transmitted to a particular application program in a specific transmission format. For example, one department in a company may use a particular application program which must receive information using a particular character as a delimiter and other departments may require the information in a different format using different delimiters.

Another problem, particularly for small businesses, is that current systems can not efficiently accommodate the inputting of information from a diversity of hard copy documents. A large business which receives many forms in the same

2

format can afford a system which inputs a high volume of information in that format into memory. For example, it is cost-effective for a bank which processes hundreds of thousands of checks a month to buy a dedicated machine which can read information off of checks having a rigidly defined, or fixed, format. However, as the diversity of forms received by a business increases relative to the number of forms that must be processed, it becomes less cost-effective to design a dedicated machine for processing each type of form format. This problem is particularly significant in small businesses which may, for example, receive fifty invoices a month, all in different, non-fixed, formats. It is frequently not cost-effective for a small business to design dedicated systems for inputting information in each of these various formats. This leaves a small business with no other practical alternative than to manually input the information off of each invoice each month.

SUMMARY OF THE INVENTION

It is an object of the invention, therefore, to provide an application program interface which allows a user to select specific portions of information extracted from a diversity of hard copy documents and allows the user to direct portions of this information to several different users in accordance with the needs of the particular user.

It is also an object of the invention to provide a cost-effective system for inputting hard copy documents which can accommodate hard copy documents in a diversity of formats.

It is another object of the invention to provide an application program interface which allows a user to put information, which is to be transmitted, into a particular transmission format, based upon the needs of the receiver of the information.

It is a further object of the invention to provide an application program interface which will allow the extraction, selection, formatting, routing, and storage of information from a hard copy document in a comprehensive manner such that the hard copy document itself need not be retained.

It is another object of the invention to provide a system which reduces the amount of manual labor required to process information originating from a hard copy document.

A further object of the invention is to reduce the time required to process information originating from a hard copy document so that a higher volume of transactions involving hard copy documents can be processed.

The invention provides an application program interface which inputs a diversity of hard copy documents using an automated digitizing unit and which stores information from the hard copy documents in a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which define portions of the stored document information required by a particular application unit. Selected stored document information is then formatted into the transmission format used by the particular application program based on transmission format instructions. The transmission formatted selected stored document information is then transmitted to the particular application program. The hard copy documents may contain textual information or image information or both.

The interface operates in three different modes.

In a first mode, the interface extracts all of the information from hard copy documents and stores this information in memory. Parsing of various portions of the extracted information is performed in accordance with content instructions.

US 7,075,673 B2

3

In a second mode, the user operates interactively with the interface by use of a display and an input device, such as a mouse. In this second mode, a hard copy document is inputted and displayed on the display. The interface then prompts the user to identify the location of various information. For example, the interface can ask the user to identify the location or address information on the hard copy document. In response, the user positions the mouse to identify address information using a cursor. The identified information is then stored as address information in memory. Subsequently, the interface again prompts the user to identify other pieces of information, which are then stored in the appropriate locations in memory. This process proceeds until all of the information which is desired to be extracted off of the hard copy document is stored in memory.

10 In a third mode of operation, selected portions of information are extracted off of hard copy documents in accordance with predetermined location information which has been specified by the user. For example, the user can define a template which specifies the location of information on hard copy documents. Templates can be formed in conjunction with second mode operation. Alternatively, the user can instruct the interface to search hard copy documents for a particular character or symbol, located on the hard copy documents. The information desired to be extracted off of the hard copy documents is specified relative to the location of this character or symbol.

The interface can also prompt or receive from an applications program or another information processing system, required information, content instructions, and format instructions.

Other objects, features, and advantages of the invention will be apparent from the following detailed description of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be described in further detail below with reference to the accompanying drawings, in which:

FIG. 1 illustrates hardware for implementing a preferred embodiment of the instant invention;

FIG. 2 illustrates an example of a hard copy document containing information to be processed by the instant invention;

FIGS. 3A and 3B are enlarged views of the computer of FIG. 1 used to explain how the invention interactively prompts a user to identify information;

FIG. 4 is an overall data flow diagram for the FIG. 1 preferred embodiment;

FIG. 5 is a detailed input data flow diagram for the FIG. 1 preferred embodiment;

FIG. 6 is a detailed information processing data flow diagram for the FIG. 1 preferred embodiment;

FIG. 7 is a more detailed information processing data flow diagram for the maintain library module of FIG. 6;

FIG. 8 is a more detailed information processing data flow diagram for the maintain definitions module of FIG. 6;

FIG. 9 is a more detailed information processing data flow diagram for the process document module of FIG. 6;

FIG. 10 is a detailed output data flow diagram for the FIG. 1 preferred embodiment;

FIG. 11 lists data corresponding to the hard copy document of FIG. 2;

FIGS. 12A, 12B, and 12C illustrate examples of data which can be selected from the extracted data of FIG. 11 in accordance with content instructions;

4

FIGS. 13A, 13B, and 13C illustrate examples of the data of FIGS. 12A, 12B, and 12C formatted in accordance with various transmission format instructions to form input files; and

5 FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Hardware

The invention provides an interface between information 15 originating from a hard copy document and a computer application unit which uses the information. The computer application unit can be a particular computer application program or a device which is controlled in accordance with instructions or information from the hard copy document. 20 The invention also allows storing a copy of the hard copy document in a memory and retrieving the copy of the hard copy document. By providing a comprehensive and integrated system which can accommodate almost all of the possible uses of information contained on a hard copy 25 document, the instant invention allows for a paperless office.

The invention includes hardware and software necessary to extract, retrieve, and process information from the hard copy document. A copy of the actual image of the hard copy document is stored in memory. Textual information 30 extracted from the hard copy document is also stored in memory. Textual information is information, such as alphanumeric characters, which is recognized on the hard copy document and which is stored in a form which corresponds to the particular recognized character. For example, the 35 extracted characters can be stored in the ASCII format in an electronic memory.

The user can have all of the information extracted from the hard copy document and stored in memory. Alternatively, the interface can interactively prompt the user to identify specific pieces of information for storage. The interface can also extract specific pieces of information using a predefined template. The interface can also prompt or receive from another information processing system or an applications program desired information, content instructions, and format instructions.

45 The instant invention also provides for parsing information extracted from the hard copy document and for directing this parsed information to specific users or application programs as an input file.

The invention also permits the user to define the transmission format of the input file for a particular computer application unit.

FIG. 1 illustrates hardware for implementing a preferred embodiment of a hard copy document to application program interface according to the instant invention. The interface 200 processes information extracted off of hard copy document 100 and provides information to application units 270 in a form required by each particular application unit. The interface extracts information off of a hard copy document 100 utilizing a scanner 210. The scanner 210 can be any type of scanner which extracts information off of hard copy documents, for example, an Optical Reader.

55 The scanned information is stored in a scanner memory 220 or in main memory 250, as will be described in greater detail below. If main memory 250 or another memory is, 60 available to store the scanned information, then scanner memory 220 can be omitted.

US 7,075,673 B2

5

The information from scanner memory 220 or main memory 250 is transmitted to computer 230. In the preferred embodiment, computer 230 includes a display 232, a keyboard 234, and a mouse 236. The display 232 displays an image of the hard copy document itself and/or information necessary to process the information extracted off of the hard copy document.

The computer 230 is used to select portions of the stored document information contained in memory in accordance with content instructions which define portions of the stored document information required by an application unit. These content instructions may be provided by the application program. Alternatively, the content instructions can be inputted via an input device such as a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

The computer 230 is also used to format selected stored document information into the transmission format used by an application unit based on transmission format instructions. The transmission format instructions may be provided by the application program. Alternatively, the transmission format instructions can be inputted via a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

Thus, the computer 230 is used to generate an input file for a particular application unit. The computer 230 is connected to scanner memory 220, main, or permanent, memory 250, a printer 260, and application units 270, via a bus 240. Although FIG. 1 illustrates use of a bus to connect components together, it is understood that any routing or connecting link, implemented in hardware or software or both, can be employed instead of, or in addition to, a bus. Instructions to or in the computer 230 control the main memory 250, the printer 260, the application units 270, and the bus 240. Instructions to or in computer 230 can also control exchanges of information with scanner memory 220.

When the computer 230 generates an input file for a particular document, the computer 230 can send this input file directly to an application unit or can store this input file in the main memory 250 until required by an application unit. The main memory 250 may also optionally store a copy of the image information for the hard copy document and the textual information for the hard copy document. Thus, the image information and textual information from the hard copy document can be retrieved and printed out on printer 260. In addition, image and textual information stored in scanner memory 220 or in main memory 250 can be used to form additional input files at the time of input or at a later time, based on content instructions and transmission format instructions. Thus, the invention can, at the discretion of the user, eliminate the need to retain copies of hard copy documents, permitting a paperless office.

The application units 270 include particular application programs and devices which are controlled in accordance with information contained on hard copy document 100.

FIG. 2 illustrates an example of a hard copy document 100 which contains information to be processed by the instant invention. The document illustrated in FIG. 2 is a bill from XYZ Corporation to customer ABC Corporation. FIG. 2 is only an example of a type of document that can be processed by the instant invention.

In a first operational mode, the scanner 210 stores all of the information extracted off of hard copy document 100 in the scanner memory 220 or, alternatively, in main memory 250. The extracted information is stored in two forms. The actual image of the hard copy document 100 is stored as image information in the scanner memory 220. In addition, the scanner memory 220 stores textual information recog-

6

nized on the hard copy document 100 by, for example, employing standard character recognition software. In the preferred embodiment, the textual information is stored in ASCII format. The scanner memory 220 can be, for example, an electronic, magnetic, or optical memory.

FIG. 3A illustrates an enlarged view of the computer 230 of FIG. 1. This view will be used to describe a second mode of operation. In this second mode of operation, the hard copy document 100 is scanned and a copy of the document 100 is displayed on display 232 of computer 230, based on the contents of information temporarily stored in scanner memory 220. After the document is displayed on display 232, the computer 230 interactively prompts the user to identify the location of specific pieces of information on the hard copy document. In the FIG. 3A illustration, this prompt message is indicated as the message beginning with the arrow.

For example, the prompt message can ask the user to identify the location of account number information on the hard copy document. The user then uses an input device, such as keyboard 234 or mouse 236 or a touch screen, notepad, voice recognition device, or other input device to position a cursor on the display to identify the location of the information requested by the prompt message. For example, the cursor could be used to define a block (which could be highlighted) containing the requested information, followed by a mouse "enter" click. In this example, the user would move the mouse to identify the location of the account number information contained on the hard copy document 100. The computer 230 then stores the information which has been identified by the user as account number information in the appropriate address or subfile or as the appropriate variable or parameter in memory. The computer then prompts the user to identify the location of other information on the hard copy document, such as, statement date information. The process proceeds until all of the desired information has been stored into the appropriate locations in memory.

FIG. 3B illustrates a variation of the second mode for interactively prompting the user for information. In FIG. 3B, the display is split into two portions. A left-hand portion 232L displays the image of the hard copy document and a right-hand portion 232R displays the required application program information. For example, in FIG. 3B, portion 232R displays a spreadsheet used by an application program. While observing the split display, the user can input instructions to associate specific pieces of information on the hard copy document (for example, the vendor name indicated by the mouse arrow 232A) with particular subfiles in memory (for example, the vendor field next to which the cursor 232C appears), using a mouse or other input device(s) or both. The split display also allows the user to generate content format instructions while observing the information required for a particular application program on the right-hand portion.

These second modes of operation are efficient for small businesses which receive a small number of a wide variety of invoices, since the user does not necessarily have to store all of the information that appears on the hard copy document. A further advantage is that data input is quicker, easier, and more accurate than with previous keyboard methodology. In addition, by specifying the location on the hard copy document of information, the user may optionally create a template, to be described in further detail below, for each different type of invoice. This template is stored for future use when another hard copy document in the same format is received.

US 7,075,673 B2

7

More specifically, instructions from computer 230 can direct the scanner 210 and scanner memory 220, and/or main memory 250, to scan and/or store only specific portions of hard copy document 100. After the interactive prompts required to obtain information for a desired application program, the unused information stored in scanner memory 220 or 250 can be erased. Further, scanning of a second identical document can be limited to only those portions of the document which contain needed information.

More specifically, in FIG. 2, the lines 10 drawn around certain portions of the document represent the areas which the user has previously identified as the portions of a document to be extracted by the scanner 210 and stored in scanner memory 220 and/or main memory 250. Since the logo 20 and the message 30 have not been identified as an area to be scanned and stored, these areas are not scanned and stored in subsequent documents. Since the user has previously associated each of the areas 10 with a specific subfile of information, e.g., the account number, the scanned information is stored in memory locations corresponding to that subfile.

Data Processing

FIGS. 4-10 illustrate the flow of data in the FIG. 1 preferred embodiment. FIG. 4 illustrates the overall data flow for the FIG. 1 preferred embodiment. The preferred embodiment includes an input process module 1.0, an information processing module 2.0, and an output processing module 3.0. The information processing module 2.0 is equipped to receive instructions from and transmit information to a user. The information processing module 2.0 can also transmit to and receive information from a remote external device through communication interface 4.0. Input process module 1.0 and output processing module 3.0 can also access communication interface 4.0. A module is implemented in hardware, software, or a combination of hardware and software. The specific implementation for a particular business application depends upon a variety of factors, for example, the relative costs of hardware and software implemented systems, the frequency with which a user will want to expand or modify the system, and the like.

FIG. 5 is a more detailed diagram of the input process module 1.0 of FIG. 4. The input process module 1.0 includes a character input module 1.1, an image input module 1.2, and, in the preferred embodiment, a character recognition device 1.3. The character input module inputs textual information, such as alphanumeric characters, from an input device such as keyboard 234. The image input module 1.2 inputs image information, for example, a digitized image of the actual appearance of hard copy document 100. Textual information can include textual input from an input device such as keyboard 234 and textual information extracted from the document by character recognition device 1.3. Both types of information comprise an input document which is transmitted to information processing module 2.0. In the FIG. 1 preferred embodiment, the processing performed by input process module 1.0 occurs in scanner memory 220, computer 230, and main memory 250.

FIG. 6 illustrates information processing data flow for the FIG. 1 preferred embodiment, that is, FIG. 6 illustrates data flow in the information processing module 2.0.

The information processing module 2.0 includes a maintain library module 2.1, to be described in further detail below in conjunction with FIG. 7, a maintain definitions module 2.2, to be described in further detail below in

8

conjunction with FIG. 8, and a process document module 2.3 to be described in further detail below in conjunction with FIG. 9.

The information processing module 2.0 is the module which coordinates and drives the entire system. In the preferred embodiment, the information processing module 2.0 is implemented primarily by computer 230.

FIG. 7 illustrates information processing data flow in the maintain library module 2.1. The maintain library module

10 2.1 maintains a library of image information, for example, a digitized image representing the actual appearance of the hard copy document, and textual information of the hard copy documents for reference during processing. This library can be incorporated within scanner memory 220, 15 main memory 250, or another independent memory, for example, a RAM disk. The maintain library module 2.1 includes a store document module 2.1.1, a correct errors module 2.1.2, a retrieve document module 2.1.3, and a document file 2.1.4. These modules operate collectively to 20 store, retrieve, and correct document information.

The store document module 2.1.1, prior to routing the document to the document file 2.1.4, may provide information on recognition errors which may have occurred while inputting the document. For example, the store document 25 module 2.1.1 identifies that a character contained in hard copy document 100 was not recognized. The store document module 2.1.1 also optionally causes a copy of the document and its parsing to be displayed on the display 232 for confirmation by the user. The user may utilize this opportunity to identify any errors in the displayed document and, 30 in conjunction with the correct errors module 2.1.2, to revise the document's parsing, if necessary, prior to storage of the document in memory. The module 2.1.1 also provides a facility for the user to name a particular hard copy document 35 for cataloging, storage, and retrieval purposes. After the document is named, the store document module 2.1.1 stores copies of the document in the document file 2.1.4.

The correct errors module 2.1.2 processes instructions 40 from the user to correct errors identified by the store document module 2.1.1 and errors that have been spotted by the user during the confirmation process.

The retrieve document module 2.1.3 permits the user to 45 retrieve a copy of a document previously stored in the document file 2.1.4. As described above, long-term storage is provided by main memory 250, if necessary.

FIG. 8 illustrates a more detailed information processing data flow diagram for the maintain definitions module 2.2 of FIG. 6. The maintain definitions module 2.2 allows the user to define system and document parameters and maintains the 50 definitions of these system and document parameters. The maintain definitions module 2.2 includes a define template module 2.2.1 which allows the user to specify the location of information on the document. This information provided by the user defines a template which is used to extract 55 information off the document and to associate the extracted information with a particular variable or subfile. These templates are illustrated by boxes 10 in the FIG. 2 example of a hard copy document. The maintain definitions module 2.2 can also access templates previously defined by the user 60 and stored in main memory 250. Templates can also be provided as part of software packages developed by program developers.

The maintain definitions module 2.2 also includes a define 65 relationships module 2.2.2. The define relationships module 2.2.2 allows the user to define data relationships, or logical relationships, between pieces of information extracted from the hard copy documents. These pieces of information are

US 7,075,673 B2

9

then used to generate an input file for a selected computer application unit. The user defines these relationships by content instructions. Alternatively, content instructions to define relationships can be provided by application software. If the user provides these content instructions, the content instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. Examples of content instructions, data, and logical relationships will be described in further detail in conjunction with FIGS. 11 and 12A, 12B, and 12C.

The maintain definitions module 2.2 also includes a define format module 2.2.3. The define format module 2.2.3 allows the user to define transmission formats for an input file which is then transmitted to a selected computer application unit. Selection of the transmission format of the input file is accomplished by the user through use of transmission format instructions. Alternatively, the applications software itself can generate its own transmission format instructions. When the user must specify transmission format instructions, the transmission format instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. A further description of various transmission formats will be provided below in conjunction with FIGS. 12A, 12B, 12C, 13A, 13B, and 13C.

A select definitions module 2.2.4 is also included in the maintain definitions module 2.2. The select definitions module 2.2.4 allows the user to store and select a set of definitions to be used for processing the document. The definitions identify pieces of information on the document by, for example, absolute location, variable location, or relative location, or by proximity to key words and/or symbols. These definitions are described in further detail below by way of an illustrative example.

FIG. 9 illustrates a more detailed information processing data flow diagram for the process document module 2.3. The process document module 2.3 processes the document after the document has been stored in the system. The process document module 2.3 gathers the appropriate information which has been stored, and creates input file(s) 2.3.3 for the selected application unit. The process document module 2.3 then transmits the input file(s) via bus 240 and/or communication interface 4.0 to an application unit 270, an output device such as printer 260, or to main memory 250.

The process document module 2.3 includes an extract data module 2.3.1. This module extracts data off of the document in accordance with the user's instructions, for example, the user-defined template, or through the interactive mode.

The process document module 2.3 also includes a preapplication process module 2.3.2 which gathers and associates information extracted from the document in accordance with content instructions. This module prompts the user for any additional information required to satisfy the relationships defined by the content instructions. The preapplication process module 2.3.2 also places the selected information into the transmission format defined by the transmission format instructions.

The preapplication process module 2.3.2 also generates the input file 2.3.3 for the selected application in accordance with the appropriate instructions. The input file 2.3.3 is then transmitted to bus 240 and/or communication interface 4.0 for transmission to a particular application unit 270.

FIG. 10 illustrates a detailed output data flow diagram for output module 3.0. Output module 3.0 outputs a textual and/or image copy of the document. In the FIG. 1 preferred embodiment, output module 3.0 is implemented by printer 260, associated software, and associated interface circuitry.

10

Operation

Examples of operation of a preferred embodiment will now be described.

5 The user enters the system by providing instructions to the information processing module 2.0. The user then instructs the information processing module 2.0 to conduct maintain library processing, maintain definitions processing, or process document processing.

10 If the user selects maintain library processing, the user then provides instructions to maintain or modify the document library through the maintain library module 2.1. For example, the user can direct the inputting and storage of a hard copy document 100 or can retrieve and output a document. The user requests inputting of a document through the store document module 2.1.1. The system then prompts the user to specify a storage location for the inputted document. The document is then read-in by the input process module 1.0. A textual copy and/or an image copy are stored into the document file 2.1.4. Errors which have occurred during inputting are identified and corrected by the correct errors module 2.1.2 and the user. The corrections are reflected in the document information stored in document file 2.1.4.

15 20 25 30 35 The retrieve document module 2.1.3 is used to retrieve and output a document. The system prompts the user to specify the storage location of a document and the type of document copy, for example, a textual or an image copy, to be outputted. The document is then outputted by the output process module 3.0.

40 45 50 55 60 65 If the user initially selected maintain definitions processing, the user would instruct the system to maintain and/or modify parameter definitions through the maintain definitions module 2.2. For example, the user can define and maintain a document template for extracting selected portions of information off of the hard copy document. The user can use the template to extract selected portions of information off of the hard copy document when the document is originally inputted, or alternatively, the user can use the template to identify selected portions of information for extraction off of an image copy of the document. In creating the template, the user identifies pieces of information on the document to be extracted and assigns a variable name, or subfile, to each piece of data.

The location of data to be extracted can be defined in a number of ways other than by use of a template. For example, the user can designate the absolute location of information on the document with respect to a grid overlaid on the document, e.g., always on line 3, starting in column 1. The user can also identify information by specifying the relative location of information to be extracted, e.g., always two lines below the piece of data named "salutation", starting in column 3. The user can also specify the location of information to be extracted by variable location specification. For example, if the hard copy document is a letter, the module would conduct a key word search for the term "Dear Sir:". Wherever this term "Dear Sir:" is located, this piece of data would be associated with the variable specified by the user, for example, the variable "salutation." In addition, a defined set of conventional symbols can be used to signify certain recurring data items for the convenience of users of the instant invention. For example, a "@" symbol can be used to delineate the vendor name as follows: "@XYZ Corporation@". Other examples of the use of symbols to delineate information will be described with reference to FIG. 14.

US 7,075,673 B2

11

The maintains definition module 2.2 is also used to maintain data relationships in accordance with content instructions and to maintain input file formats in accordance with transmission format instructions. Relationships are defined and maintained between pieces of data, specified by, for example, the names of variables, through the define relationships module 2.2.2. The names of pieces of data on the document are retrieved by, for example, the define template module 2.2.1, and are passed to the define relationships module 2.2.2. The user may then provide any additional pieces of data needed to generate an input file for a particular application program or unit, such as an input file line number. The user, the applications software, and/or instructions previously stored in memory then establishes the contents of the input file by defining relationships between pieces of data using content instructions. Specific examples of content instructions will be discussed below in conjunction with FIGS. 11, 12A, 12B, 12C, 13A, 13B, and 13C.

The user and/or the applications software defines and maintains the transmission format of the input file to be used by a particular application program or unit through the define format module 2.2.3 in accordance with transmission format instructions. This is accomplished by defining the parameters to be used by the preapplication process module 2.3.2 in generating an input file. Parameters which would typically be required to generate an input file would include the character type, e.g., text or pixel; delimiters used between pieces of data, e.g., a slash or a semicolon; end of line characters, e.g., a carriage return or a line feed; and end of file characters. Examples of transmission formats will be described in further detail below in conjunction with FIGS. 11, 12A, 12B, 12C, 13A, 13B, and 13C.

If the user initially selected process document processing, the interface will then proceed to process the document through use of the process document module 2.3. For example, the user can extract specific portions of data from an image copy of a document, can generate an input file for transmission to an application program, or can directly process information interactively with an application program.

If the user desires to extract specific portions of data from an image copy of a hard copy document which has already been stored in memory, is the user uses the extract data module 2.3.1 to identify a document to be processed. The document is then retrieved by the retrieve document module 2.1.3 and passed to the extract data module 2.3.1. The user can also select parameter definitions through the select definitions module 2.2.4.

The selected document template or parameter definition is passed to the extract data module 2.3.1. The extract data module 2.3.1 extracts pieces of data from the image copy of the document, as defined by the document template definition or the parameter definitions or both. This document data is then passed to preapplication process module 2.3.2.

The interface generates input file(s) 2.3.3 by use of the preapplication process module 2.3.2. The selected data relationship definition, as defined by the content instructions, and the selected record format definitions, as defined by the transmission format instructions, are passed to the preapplication process module 2.3.2. The preapplication process module 2.3.2 assembles the input file in accordance with the content instructions. The preapplication process module 2.3.2 also prompts the user for any additional pieces of data which need to be provided by the user. The input file is converted to the desired transmission format in accordance

12

with the transmission format instructions. This physically formatted data is then stored in the input file 2.3.3.

The user can also use an application program to process information by loading the particular application program into the computer 230 rather than by sending the input file to a remote application unit 270.

An illustrative example of the processing described above will now be described.

The user inputs instructions via keyboard 234 or another input device which indicate that the user desires to input and store a document. The computer 230 then prompts the user for the name of the document. In this example, the user desires to input the document of FIG. 2 and therefore names the document "XYZ Corp. Bill 12/01/86." The computer then prompts the user to feed the hard copy document 100 into the scanner 210. The image of the hard copy document is displayed on display 232. The computer then prompts the user to identify the account number on the document. By use of the mouse 236 or other input device to position a cursor on the display, the user indicates the location of the account number. The account number is then read-in to a subfile named "Account Number." This process proceeds until all of the desired information has been read-in and stored.

In this particular example, no errors were encountered while inputting the document. The user then directs that the document be stored for future reference in a document file.

Some time later, the user desires to retrieve and output the document and to generate input files based on information from the document. The computer 230 prompts the user for the name of the document and the type of output. The user responds with "XYZ Corp. Bill 12/01/86" for a printed textual copy. The document is then retrieved from the document file and passed to the printer 260 for printing.

In order to generate an input file for a specific application program, the user selects the option to define a document template for use when each month's XYZ Corporation bill arrives. Accordingly, the user instructs the system to display a copy of an XYZ Corporation bill on the display 232. The user then identifies pieces of data by absolute locations. That is, the user assigns specific names to information located at specific portions of the document. In this example, the user would input the following information:

Vendor-text, line 1, one line, column 1, 80 characters;

Account number-numeric, line 6, one line, column 25, 9 characters;

Statement date-date, line 9, one line, column 25, 8 characters;

Payment date-date, line 11, one line, column 25, 8 characters;

Previous balance-currency, line 7, one line, column 75, 9 characters;

New charges-currency, line 8, one line, column 75, 9 characters;

Other debits-currency, line 10, one line, column 75, 9 characters;

Finance charges-currency, line 12, one line, column 75, 9 characters;

Payments-currency, line 13, one line, column 75, 9 characters;

Other credits-currency, line 14, one line, column 75, 9 characters;

New balance-currency, line 15, one line, column 75, 9 characters.

The user also identifies data with variable locations. In this particular example, a variable location is specified as follows:

US 7,075,673 B2

13

Heading 2—line, value “Mail To.”

The identification of Heading 2 as line information means that the system will search for occurrences of the character string “Mail To:” and assign the line number which contains this character string to Heading 2.

The user also identifies data by relative locations. In this example, the user identifies the following relative location:

Mail To-text, Heading 2+1, 3 lines, column 60, 25 characters per line.

The instructions above instruct the system to assign the textual information beginning on one line after Heading 2 and continuing for 3 lines, in column 60, to the Mail To subfile.

As an alternative to inputting the actual line, column, and character numbers, the user can identify desired portions of the document by blocking, or highlighting, the desired portions using the mouse or other input device. In this case, the computer converts the highlighted portions into corresponding line, column, and character numbers.

FIG. 11 lists data corresponding to the hard copy document of FIG. 2 and the associated variable or subfile names.

Next, the user desires to define data relationships in accordance with content instructions. Examples of the type of contents which can be specified by a user are illustrated in FIGS. 12A, 12B, and 12C.

In this particular example, three separate departments of ABC Corporation require information from the XYZ Corporation bill. The first department requires vendor, account number, statement date, payment date, previous balance, new charges, debits, finance charges, payments, and new balance information. The second and third departments require mail to information and previous balance information. Each of these departments have their own application program which utilizes this information.

The user employs content instructions to designate how pieces of information, which have been extracted off of hard copy document 100, are directed to particular departments, that is, particular application programs. FIG. 12A illustrates the contents of the information to be transmitted to the first department. FIG. 12B illustrates the information to be transmitted to the second department. FIG. 12C illustrates the information to be transmitted to the third department. The content instructions, therefore, parse the information shown in FIG. 11 to various application programs, as shown by FIGS. 12A, 12B, and 12C. Content instructions can also be used to identify additional pieces of data which are required for the input files of the particular application programs. In this particular example, the specific application programs from the three departments all require numeric record number information, numeric horizontal position information, numeric vertical position information, and date received information. The horizontal and vertical position information is used by the application program to specify the location of the received information on a spreadsheet application program, in this example. The user may know in advance the content format required by each application program, that is, in this example, the location and type of information specified on the spreadsheet. The user may also employ the split display mode described with reference to FIG. 3B to generate content format instructions.

Using the content instructions, the user establishes the following contents for the input file corresponding to FIG. 12A:

Record number, horizontal position, vertical position, vendor;

Record number, horizontal position, vertical position, account number;

14

Record number, horizontal position, vertical position, statement date;

Record number, horizontal position, vertical position, date received;

5 Record number, horizontal position, vertical position, payment date;

Record number, horizontal position, vertical position, previous balance;

Record number, horizontal position, vertical position, new charges;

Record number, horizontal position, vertical position, finance charges;

Record number, horizontal position, vertical position, payments;

15 Record number, horizontal position, vertical position, new balance.

Next, transmission format instructions are employed to define the transmission format of the input file for a specific application program or unit. FIG. 13A illustrates the transmission input file corresponding to FIG. 12A. FIG. 13B illustrates the transmission input file corresponding to FIG. 12B. FIG. 13C illustrates the transmission input file corresponding to FIG. 12C. A comparison of FIGS. 12B and 12C reveals that FIGS. 12B and 12C have the same contents. However, the information illustrated in FIG. 12B is being sent to a different application program than the information in FIG. 12C. These application programs require different transmission input formats, as illustrated in FIGS. 13B and 13C. More specifically, the application program that receives the input file illustrated in FIG. 13B uses the greater than sign as a delimiter whereas the application program which receives the transmission input file shown in FIG. 13C uses a back-slash as the delimiter.

35 After the contents and the transmission format for the input file have been defined, and any additional information has been inputted, the input file is assembled and transmitted to the particular application program.

FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention. The hard copy document illustrated in FIG. 14 is first scanned and information from the hard copy document is stored into a memory. The interface 200 then identifies portions of the hard copy document corresponding to various variables by recognizing a defined set of symbols. In the FIG. 14 example, triangles delineate the mailing address, circles delineate the statement date, and squares delineate the new charges. Information from these portions of the hard copy document is stored in the corresponding memory locations or subfiles for each variable. The same set of symbols can be used to identify the same information from one document to the next. Thus, even if the physical formats of documents are not fixed from one document to the next, a diversity of hard copy documents can be processed without manually inputting data by recognition of the defined symbols.

Examples of readily available application programs are Quicken and Lotus 1,2,3 both of which are widely utilized in the business community. Quicken, for example is an 60 easy-to-utilize program for writing checks and preparing business records. Payee, amount and address information may readily be transmitted from scanner memory 220 and/or main memory 250 to the Quicken application program for check writing functions and ledger keeping purposes. Lotus is a well known spreadsheet program which may process data input into specified cells once this data is placed in conventional Lotus format.

US 7,075,673 B2

15

Thus, the instant invention provides an integrated and comprehensive system for handling information from a hard copy document, thus permitting a paperless office. In addition, the invention permits data, extracted off of a hard copy document, to be easily manipulated into various logical and transmission formats required by a particular application unit. The invention also provides a low cost system for inputting information from a wide variety of hard copy documents into a memory.

The foregoing description has been set forth merely to illustrate preferred embodiments of the invention and is not intended to be limiting. Modifications are possible without departing from the scope of the invention.

For example, letters, checks, forms, pictures, reports, music scores, film, and other types of hard copy documents can be processed by the invention for accounts payable/receivable accounting, inventory control, record keeping, budgeting, data base management, music transcription, forms processing, computerized art, survey and questionnaire processing, statistical data analysis, correspondence processing and other applications.

Other automated digitizing units can be used in addition to or as an alternative to use of the scanner 210 as an input unit. Any electrical, magnetic, or optical device which extracts information off of a hard copy document, thereby eliminating the need to manually input significant amounts of information from the hard copy document is suitable for use as an automated digitizing unit. In addition, information can be input by user responses and digital and analog signals generated from various devices, and from computer files from other computer systems. Suitable hardware for inputting data includes a keyboard, a light pen, a mouse, a touch screen, a laser scanner, a microphone, a tablet, a disk drive, a magnetic tape drive, and a modem.

The interface 200 can also output information in forms other than a hard copy of textual or image information. For example, the interface 200 can output system responses, computer files, and digital and analog signals for transmission to other computer systems or to control systems. Suitable hardware for outputting information includes a disk drive, a magnetic tape drive, a cathode ray tube, a plasma screen, a printer, a plotter, a film developer, an amplifier, and a modem.

Since modifications of the described embodiments incorporating the spirit and substance of the invention may occur to persons skilled in the art, the scope of the invention should be limited solely with respect to the appended claims and equivalents.

What is claimed is:

1. A system configurable to input information from an electronic document template displayed on a display of a first computer into a plurality of application programs operable on another computer in which said template automatically generates content instructions to extract data and to transmit said data to at least one field in at least one of said plurality of application programs according to customizable transmission format instructions which are dependent upon and compatible with said at least one application program, operating on said other computer, that is receiving said transmitted data.

2. The system as recited in claim 1 in which said system is a distributed computing system.

3. The system as recited in claim 1 wherein said content instructions automatically parse said data to said at least one field in said application program.

16

4. The system as recited in claim 1 wherein said content instructions are automatically generated using a mouse function.

5. The system as recited in claim 1 wherein said content instructions are automatically generated using speech recognition.

6. The system as recited in claim 1 wherein said content instructions are automatically generated interactively with said user.

10. 7. The system as recited in claim 1 wherein said extracted data comprises billing information.

8. The system as recited in claim 1 wherein said extracted data comprises financial information.

9. The system as recited in claim 1 wherein said extracted data comprises personal information.

15. 10. The system as recited in claim 1 wherein said extracted data comprises customer information.

20. 11. A method of doing business utilizing a computer system configurable to input information from an electronic document template displayed on a display of a first computer into a plurality of application programs operable on another computer comprising the steps of:

displaying an image of the electronic document on the first computer;

utilizing the image of the electronic document of said first computer as said template;

utilizing said template to automatically generate content instructions;

30. extracting data corresponding to said information from the image of the electronic document utilizing said content instructions;

transmitting said extracted data to at least one field in said at least one of said plurality of application programs on said other computer;

35. said transmitting being done in accordance with customizable transmission format instructions which are transmission format instructions customizable to be dependent upon and compatible with said at least one of said plurality of application programs operating on said other computer that is receiving said transmitted data.

40. 12. The method as recited in claim 11 further comprising the step of executing the application program on said other computer.

45. 13. The method as recited in claim 11 wherein said extracted data comprises billing information.

14. The method as recited in claim 11 wherein said extracted data comprises financial information.

50. 15. The method as recited in claim 11 wherein said extracted data comprises personal information.

16. The method as recited in claim 11 wherein said extracted data comprises customer information.

55. 17. An interface of a computer system configurable to input information from an electronic document template displayed on a display of a first computer into a plurality of application programs operable on another computer, said interface for inputting and transmitting information comprising:

means for displaying an image of the electronic document on the first computer;

means for utilizing the image of the electronic document of said first computer as said template;

means, utilizing said template, to automatically generate content instructions;

means for extracting data corresponding to said information from the image of the electronic document utilizing said content instructions;

US 7,075,673 B2

17

means for transmitting said extracted data to at least one field in said at least one of said plurality of application programs on an other computer, different from said first computer and in accordance with customizable transmission format instructions which are transmission format instructions customizable to be compatible with said at least one of said plurality of application programs, operating on said other computer, that is receiving said transmitted data.

18. The interface as recited in claim 17 wherein said content instructions are automatically generated interactively with said user.

19. The interface as recited in claim 17 wherein said content instructions are automatically generated using speech recognition.

20. The interface as recited in claim 17 wherein said content instructions are automatically generated using a mouse function.

21. The interface as recited in claim 17 wherein said extracted data comprises billing information.

22. The interface as recited in claim 17 wherein said extracted data comprises financial information.

23. The interface as recited in claim 17 wherein said extracted data comprises personal information.

24. The interface as recited in claim 17 wherein said extracted data comprises customer information.

25. A method utilizing a computer system configurable to input information from an electronic document template displayed on a display of a first computer into a plurality of application programs operable on another computer, said method comprising the steps of:

displaying an image of the electronic document on the first computer;

utilizing the image of the electronic document of said first computer as the template;

utilizing said template to automatically generate content instructions;

extracting data corresponding to said information from the image of the electronic document utilizing said content instructions;

transmitting said extracted data to at least one field in said at least one of said plurality of application programs operable on said other computer;

said transmitting being done in accordance with customizable transmission format instructions which are transmission format instructions customizable to be dependent upon and compatible with said at least one of said plurality of application programs, operating on said other computer that is receiving said transmitted data.

26. The method as recited in claim 25 further comprising the step of executing the application program on said other computer.

27. The method as recited in claim 25 wherein said extracted data comprises billing information.

28. The method as recited in claim 25 wherein said extracted data comprises financial information.

29. The method as recited in claim 25 wherein said extracted data comprises personal information.

30. The method as recited in claim 25 wherein said extracted data comprises customer information.

31. An application program interface for use in data processing in a computer system configurable to input information from an electronic document template displayed on a display of a first computer into a plurality of application programs operable on another computer, said interface comprising:

18

utilities for enabling automatic generation of content instructions using the electronic document template, said content instructions designating data portions of said electronic document template required by at least one of said plurality of application programs;

utilities for enabling formatting of said data portions used by said at least one of said plurality of application programs based on customizable transmission format instructions which are transmission format instructions customizable to be dependent upon and compatible with said at least one of said plurality of application programs; and

utilities for automatically enabling transmission of formatted selected data portions to said at least one of said plurality of application programs, operating on said other computer that is receiving said transmitted data.

32. The application program interface as recited in claim 31, wherein said content instructions automatically parse said data to said at least one field in said application program.

33. The application program interface as recited in claim 31, wherein said content instructions are automatically generated using a mouse function.

34. The application program interface as recited in claim 31, wherein said content instructions are automatically generated using speech recognition.

35. The application program interface as recited in claim 31, wherein said content instructions are automatically generated interactively with said user.

36. The application program interface as recited in claim 31, further comprising utilities for enabling storage of said content instructions.

37. The application program interface as recited in claim 31, wherein said interface operates in a distributed computing system.

38. The application program interface as recited in claim 31, further comprising utilities for enabling definition of said electronic document template for associating portions of said electronic document template with specific variables.

39. The application program as recited in claim 31 wherein said designated data comprises billing information.

40. The application program as recited in claim 31 wherein said designated data comprises financial information.

41. The application program as recited in claim 31 wherein said designated data comprises personal information.

42. The application program as recited in claim 31 wherein said designated data comprises customer information.

43. A method of processing information utilizing a computer system configurable to input information from an electronic document template displayed on a display of a first computer into a plurality of application programs operable on another computer, said method comprising the steps of:

displaying an image of an electronic document on the first computer for inputting information into at least one of said plurality of application programs stored on an other computer according to transmission format instructions;

processing the information by automatically generating content instructions to parse at least a portion of the image to at least one field of information required by said at least one of said plurality of application programs; and

US 7,075,673 B2

19

transferring said at least one field of information to said at least one of said plurality of application programs utilizing customizable transmission format instructions which are transmission format instructions customizable to be dependent upon and compatible with said at least one of said application programs, operating on said other computer that is receiving said transmitted data.

44. The method as recited in claim 43 wherein said information comprises billing information.

45. The method as recited in claim 43 wherein said information comprises financial information.

46. The method as recited in claim 43 wherein said information comprises personal information.

47. The method as recited in claim 43 wherein said information comprises customer information.

48. An information processing system configurable to input information from a first computer into a plurality of application programs operable on another computer in which a template document transmitted by said first computer to said other computer includes content instructions to automatically extract data and to transmit said data to at least one field in at least one of said plurality of application programs according to customizable transmission format instructions which are transmission format instructions dependent upon and customizable to be compatible with said at least one of said plurality of applications programs, operating on said other computer that is receiving said transmitted data.

20

49. The information processing system as recited in claim 48 in which said system is a distributed computing system.

50. The information processing system as recited in claim 48 wherein said content instructions automatically parse said data to said at least one field in said application program.

51. The information processing system as recited in claim 48 wherein said content instructions are automatically generated using a mouse function.

10 52. The information processing system as recited in claim 48 wherein said content instructions are automatically generated using speech recognition.

15 53. The information processing system as recited in claim 48 wherein said content instructions are automatically generated interactively with said user.

54. The method as recited in claim 48 wherein said extracted data comprises billing information.

55. The information processing system as recited in claim 48 wherein said extracted data comprises financial information.

20 56. The information processing system as recited in claim 48 wherein said extracted data comprises personal information.

25 57. The information processing system as recited in claim 48 wherein said extracted data comprises customer information.

* * * * *



US007184162B2

(12) **United States Patent**
Lech et al.

(10) **Patent No.:** **US 7,184,162 B2**
(45) **Date of Patent:** **Feb. 27, 2007**

(54) **INFORMATION PROCESSING
METHODOLOGY**

(75) Inventors: **Robert Lech**, Jackson, NJ (US);
Mitchell A. Medina, Nairobi (KE);
Catherine B. Elias, Plainsboro, NJ (US)

(73) Assignee: **Eon-Net L.P.**, Tortola (VG)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 52 days.

(21) Appl. No.: **11/108,508**

(22) Filed: **Apr. 15, 2005**

(65) **Prior Publication Data**

US 2005/0185218 A1 Aug. 25, 2005

Related U.S. Application Data

(60) Continuation of application No. 10/704,484, filed on Nov. 6, 2003, now Pat. No. 7,075,673, which is a continuation of application No. 09/458,162, filed on Dec. 9, 1999, now Pat. No. 6,683,697, which is a continuation of application No. 09/044,159, filed on Mar. 19, 1998, now Pat. No. 6,094,505, which is a continuation of application No. 08/487,150, filed on Jun. 7, 1995, now Pat. No. 5,768,416, which is a division of application No. 08/348,224, filed on Nov. 28, 1994, now Pat. No. 5,625,465, which is a continuation of application No. 08/143,135, filed on Oct. 29, 1993, now Pat. No. 5,369,508, which is a continuation of application No. 07/672,865, filed on Mar. 20, 1991, now Pat. No. 5,258,855.

(51) **Int. Cl.**
H04N 1/40 (2006.01)

(52) **U.S. Cl.** **358/1.15; 382/175**
(58) **Field of Classification Search** **358/1.15;**
382/175, 177, 180, 282, 287, 306, 317

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,013,718 A 12/1961 Shepard et al.

(Continued)

FOREIGN PATENT DOCUMENTS

EP 0 107 083 B1 7/1988

(Continued)

OTHER PUBLICATIONS

Ron Pernick, A Timeline of the First Ten Years of The Well, Well Historical timeline—the good, great place, 1995, <http://www.well.com/conf/welltale/timeline.html>.

(Continued)

Primary Examiner—Thomas D. Lee

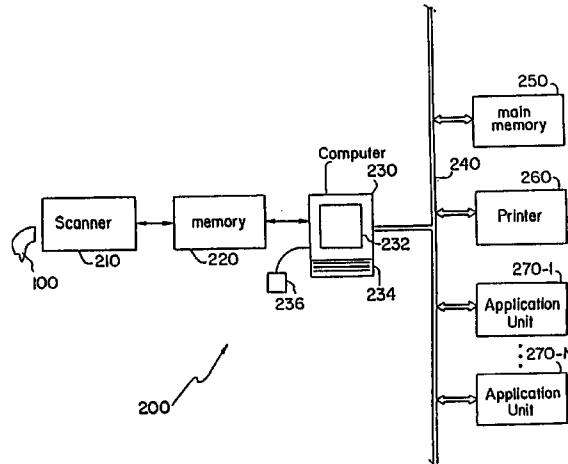
Assistant Examiner—Stephen Brinich

(74) *Attorney, Agent, or Firm*—Foley & Lardner LLP

(57) **ABSTRACT**

An information processing methodology gives rise to an application program interface which includes an automated digitizing unit, such as a scanner, which inputs information from a diversity of hard copy documents and stores information from the hard copy documents into a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which designate portions of the stored document information required by a particular application program. The selected stored document information is then placed into the transmission format required by a particular application program in accordance with transmission format instructions. After the information has been transmission formatted, the information is transmitted to the application program. In one operational mode, the interface interactively prompts the user to identify, on a display, portions of the hard copy documents containing information used in application programs or for storage.

56 Claims, 15 Drawing Sheets



US 7,184,162 B2

Page 3

Industry News, Remittance and Document Processing Today, Jan. 1989, p. 9.

Scott Beamer, *Mac OCR Takes a Big Step Forward*, MacWeek, Jun. 13, 1989.

Matthew Lake, *Strength of Character (Recognition)*, Publish, Jan. 1991, pp. 62-67.

R. David Nelson and Karen A. Hamill, *Optical Scanning at Chemical Abstracts Service for Building Computer Files From Printed Index Data*, Recognition Technologies Today, Feb. 1985, pp. 1-6, 15.

Gerald Farmer, *HNC IDEPT 9 and Recognition Enhanced Data Entry: The Cost-Cutting Approach to Automated Data Entry*, Remittance and Document Processing Today, Jan.-Feb. 1991, pp. 24-26.

David Gertler, *Automated Data Entry*, Seybold Report on Desktop Publishing, Jan. 15, 1990, pp. 3-17.

Eric Aas and Peter Davidoff, *Teaching Your Scanner to Read*, Personal Publishing, May 1990, pp. 28, 31, 33.

Phillip Robinson et al., *Character Witnesses*, MacUser, Jul. 1990, pp. 120-136.

Brita Meng, *Text Without Typing*, MacWorld, Oct. 1990, pp. 177-183.

Jim Heid, *Getting Started with Optical Character Recognition*, MacWorld, Oct. 1990, pp. 297-301.

Stanford Diehl and Howard Eglowstein, *Tame the Paper Tiger*, Byte, Apr. 1991, pp. 220-238.

Alan Joch and Rich Graham, *Voices of Experience*, Byte, Apr. 1991, pp. 239-241.

Gregory Boleslavsky and Roman Tutunikov, *The New Generation of OCR*, Inform, Jan. 1990, pp. 34-37.

Calera Recognition Systems, Inc., TopScan Professional User's Guide: Complete Document Recognition for PCs and Compatibles (1989).

Calera Recognition Systems, Inc., TopScan Professional Installation Notes for Scanners, Fax Cards, and System Configuration (1989).

Calera Recognition Systems, Inc., TopScan Professional Troubleshooting Guide (1989).

Xerox Imaging Systems, Inc., Datacopy AccuText User's Guide (1989).

Invoice from Corporate Intelligence Corporation to Workman, Nydegger & Seeley, dated Sep. 21, 1999.

Examiner Interview Summary Record, Jan. 25, 1996, and related Amendment, for Application U.S. Appl. No. 08/097,131.

Jane B. Newman, Formstar Ad; "Stack the Facts, Not the Forms"—May 24, 1987; "Stack and Send Just the Facts—and Improve the Efficiency of your Forms Application".

TeleImage Systems Document and Image Database Systems User's Manual-Ramat Gan, Israel; Table of Contents and pp. 2-1 through 5-20.

Form Out! Programmer's Manual; TeleImage Systems, Ramat Gan Israel; published Feb. 1991 pp. I, II, III, IV, V, VI, VII; 1.1-6.29 and A.1-G.11.

* cited by examiner

U.S. Patent

Feb. 27, 2007

Sheet 1 of 15

US 7,184,162 B2

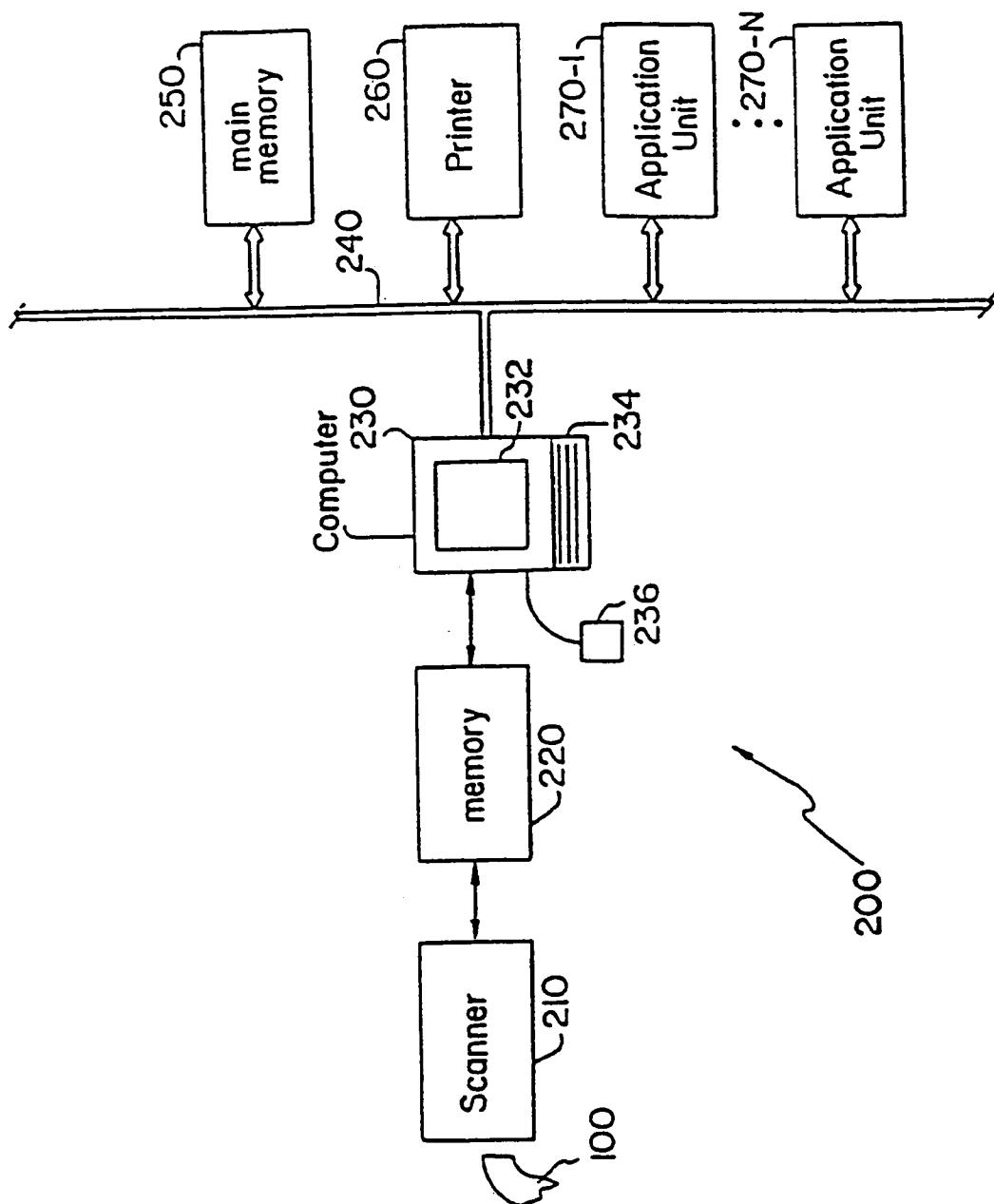


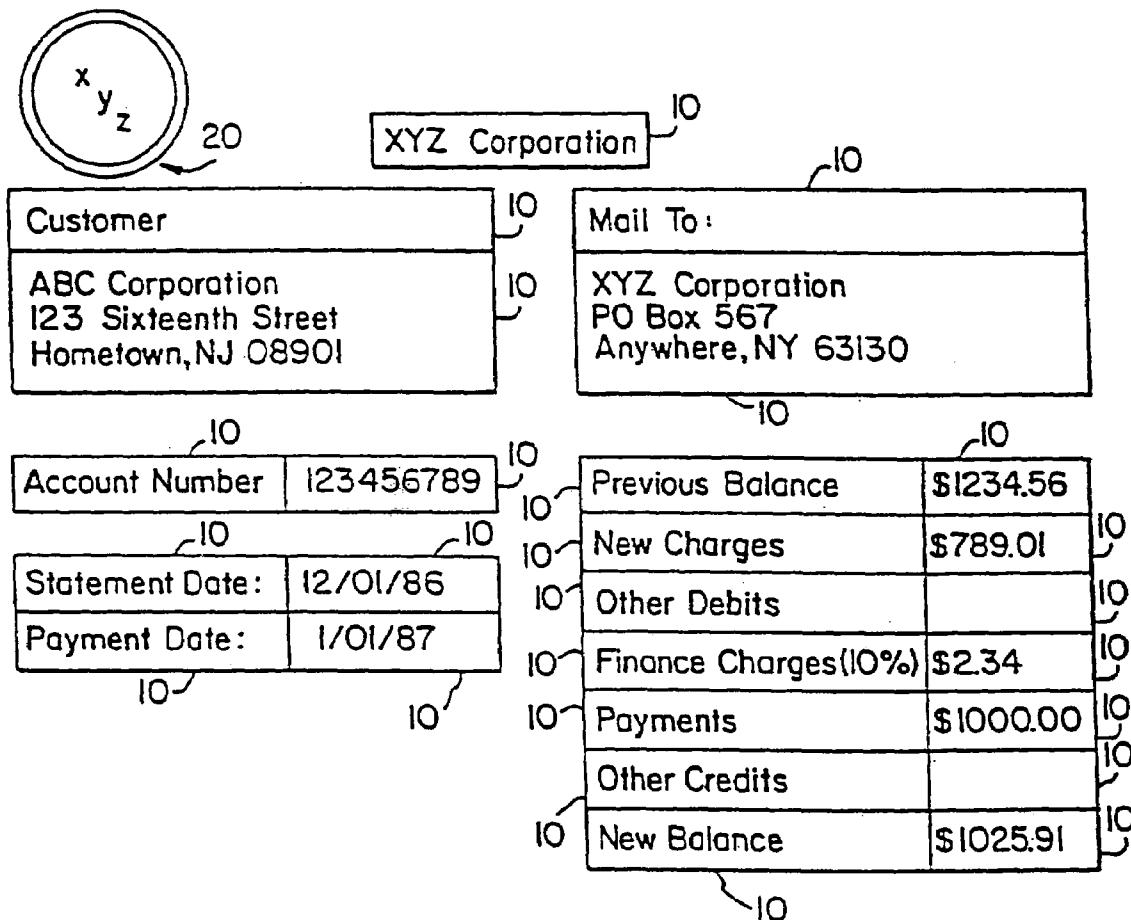
FIG. I

U.S. Patent

Feb. 27, 2007

Sheet 2 of 15

US 7,184,162 B2



Winter Sale In Effect Throughout January!

30

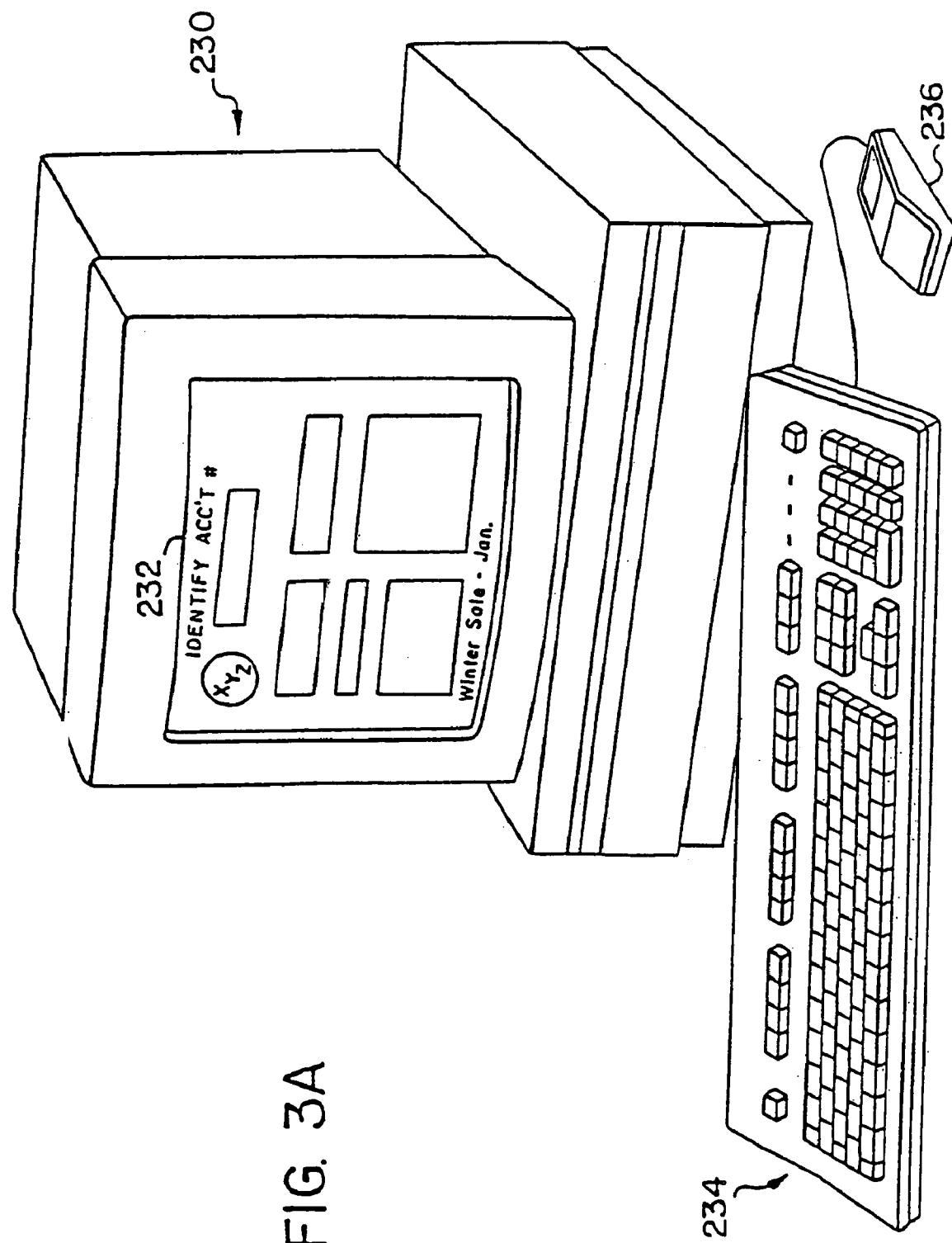
FIG. 2

U.S. Patent

Feb. 27, 2007

Sheet 3 of 15

US 7,184,162 B2



U.S. Patent

Feb. 27, 2007

Sheet 4 of 15

US 7,184,162 B2

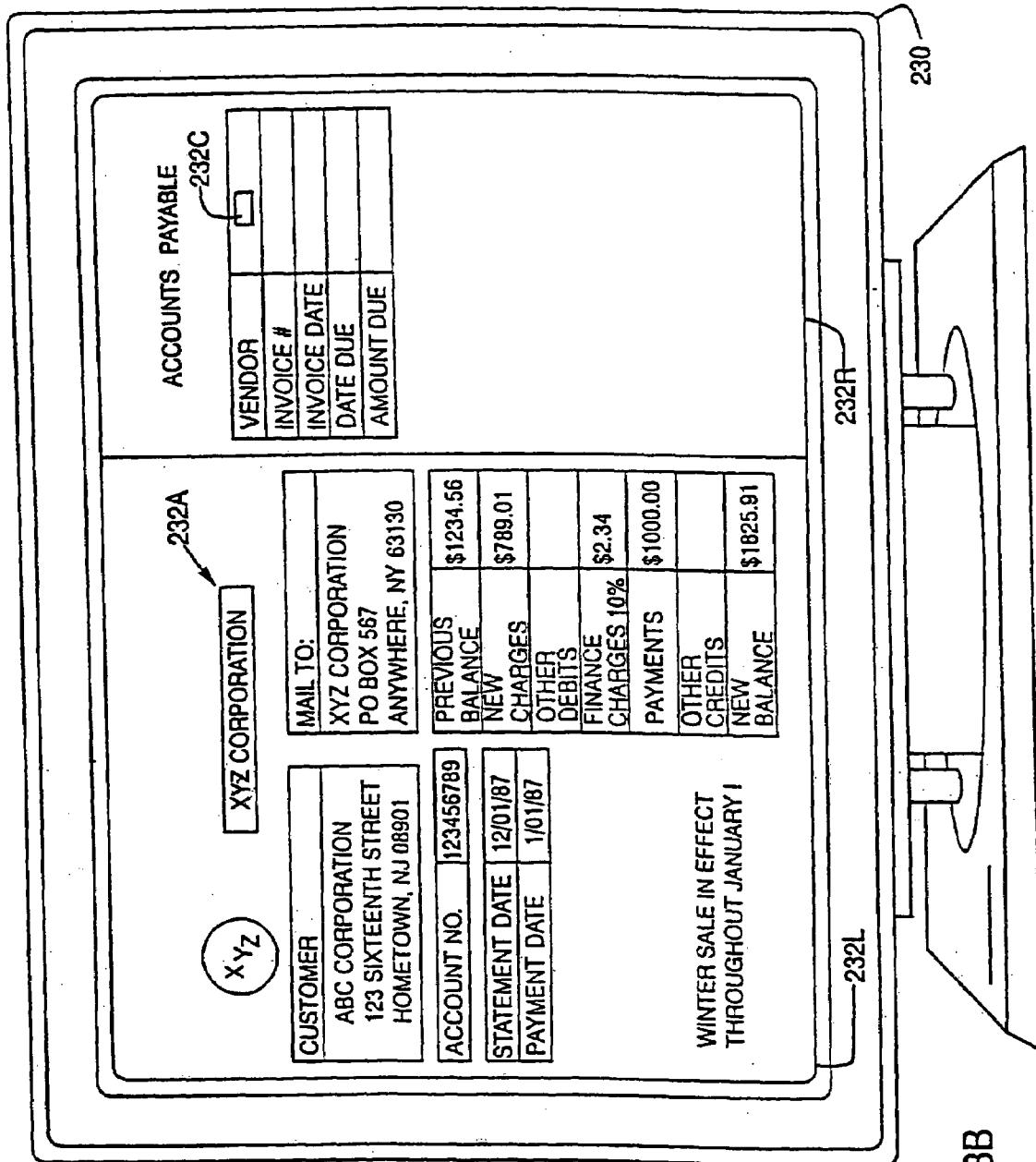


FIG. 3B

U.S. Patent

Feb. 27, 2007

Sheet 5 of 15

US 7,184,162 B2

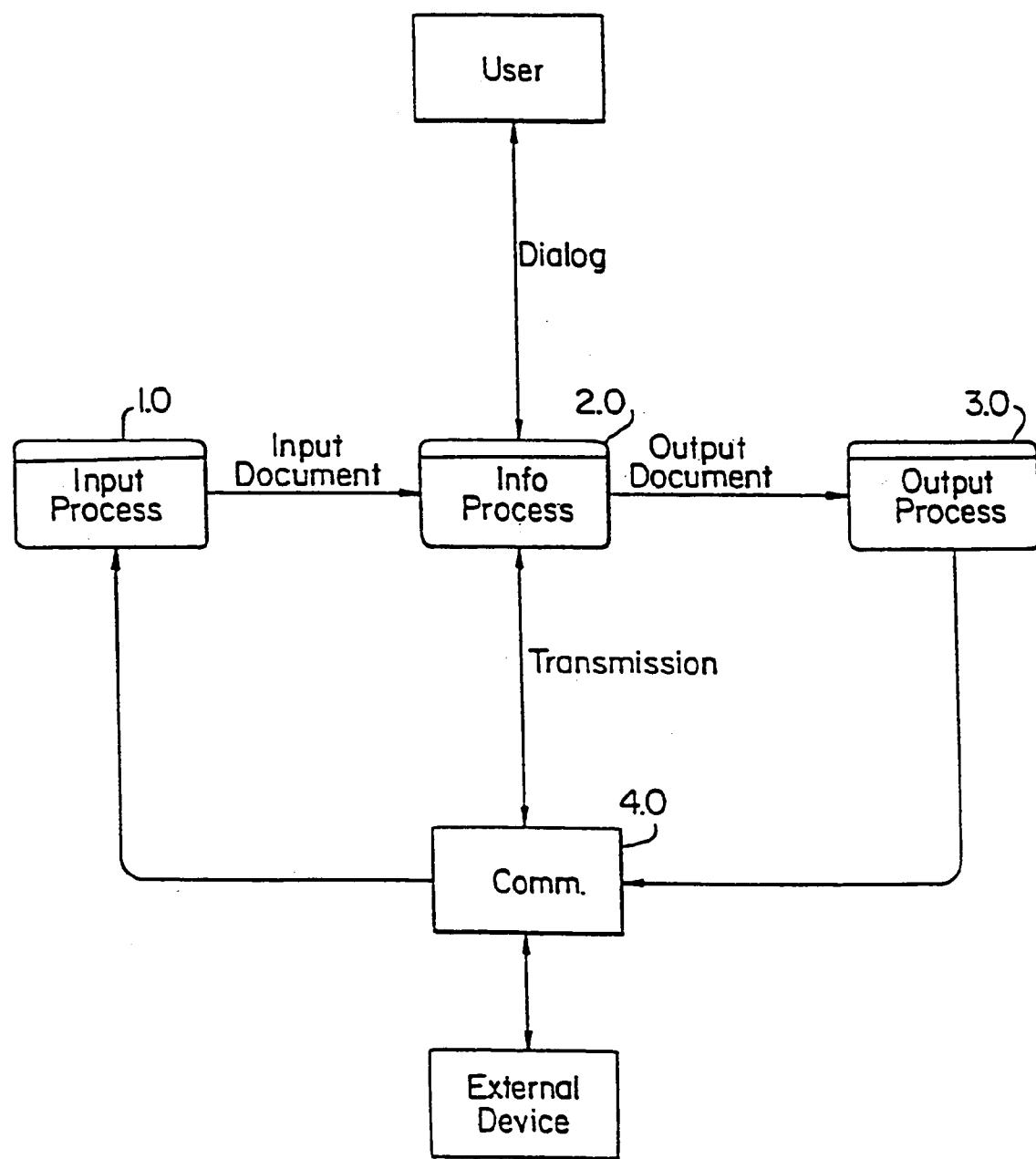


FIG. 4

U.S. Patent

Feb. 27, 2007

Sheet 6 of 15

US 7,184,162 B2

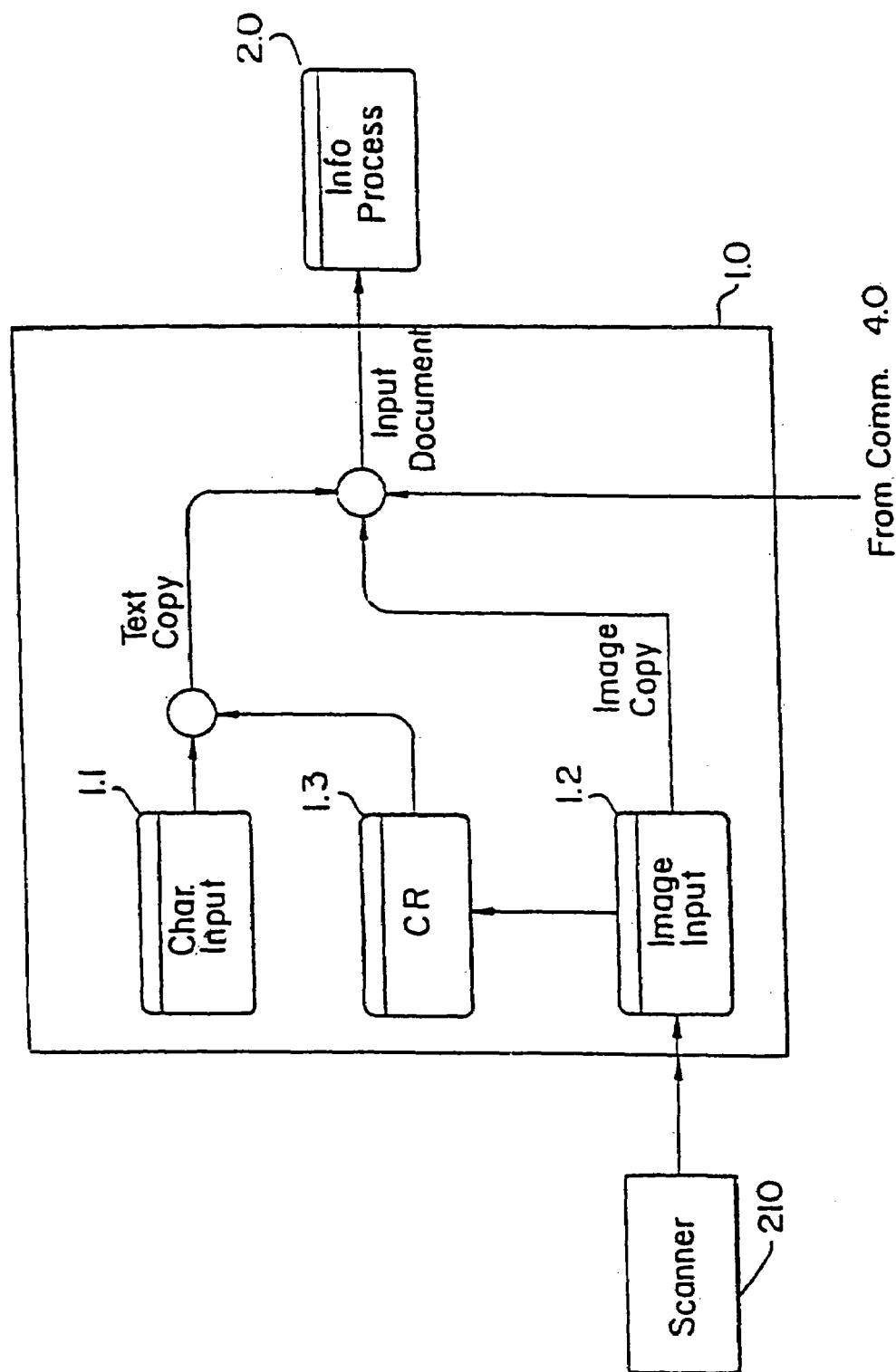


FIG. 5

U.S. Patent

Feb. 27, 2007

Sheet 7 of 15

US 7,184,162 B2

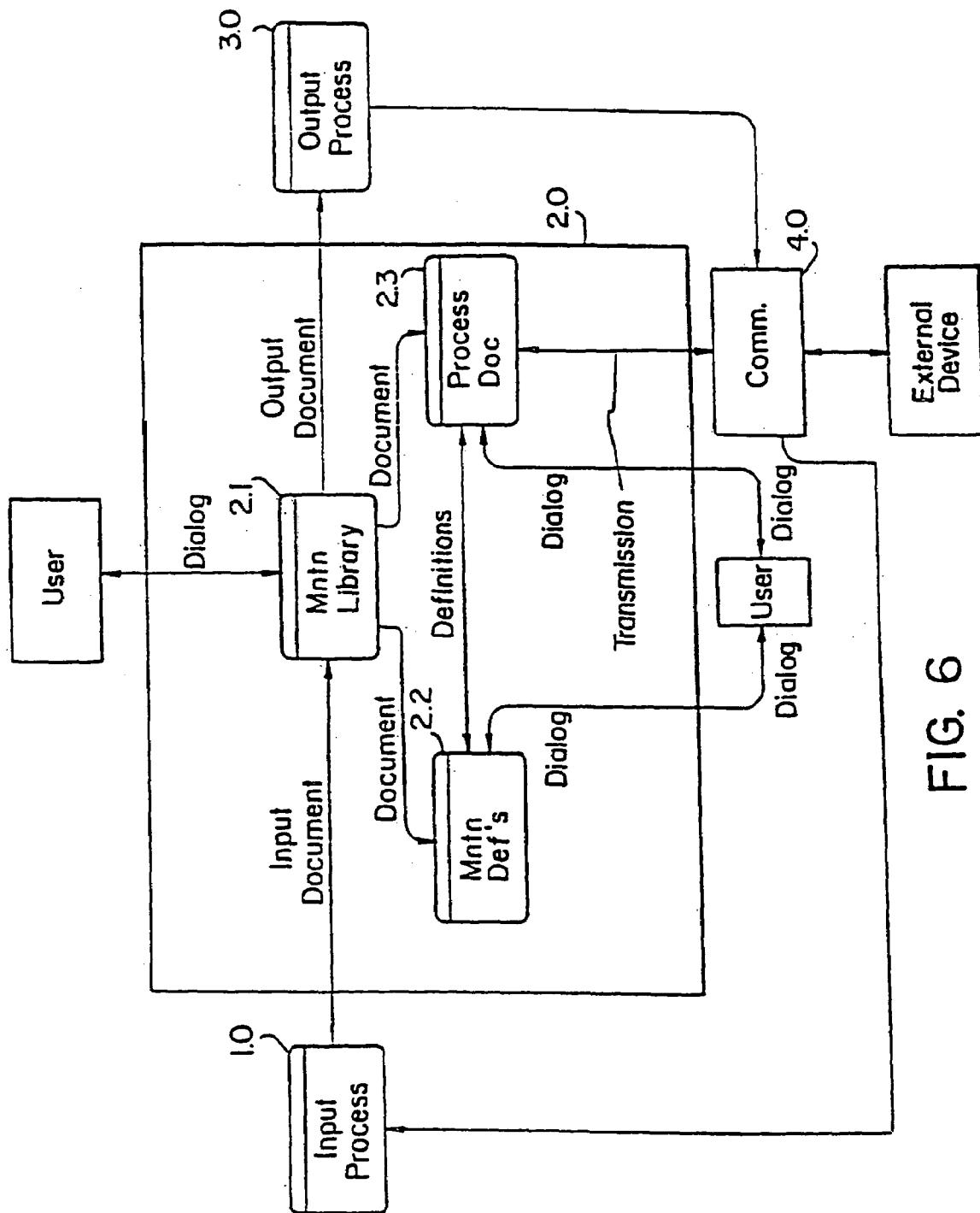


FIG. 6

U.S. Patent

Feb. 27, 2007

Sheet 8 of 15

US 7,184,162 B2

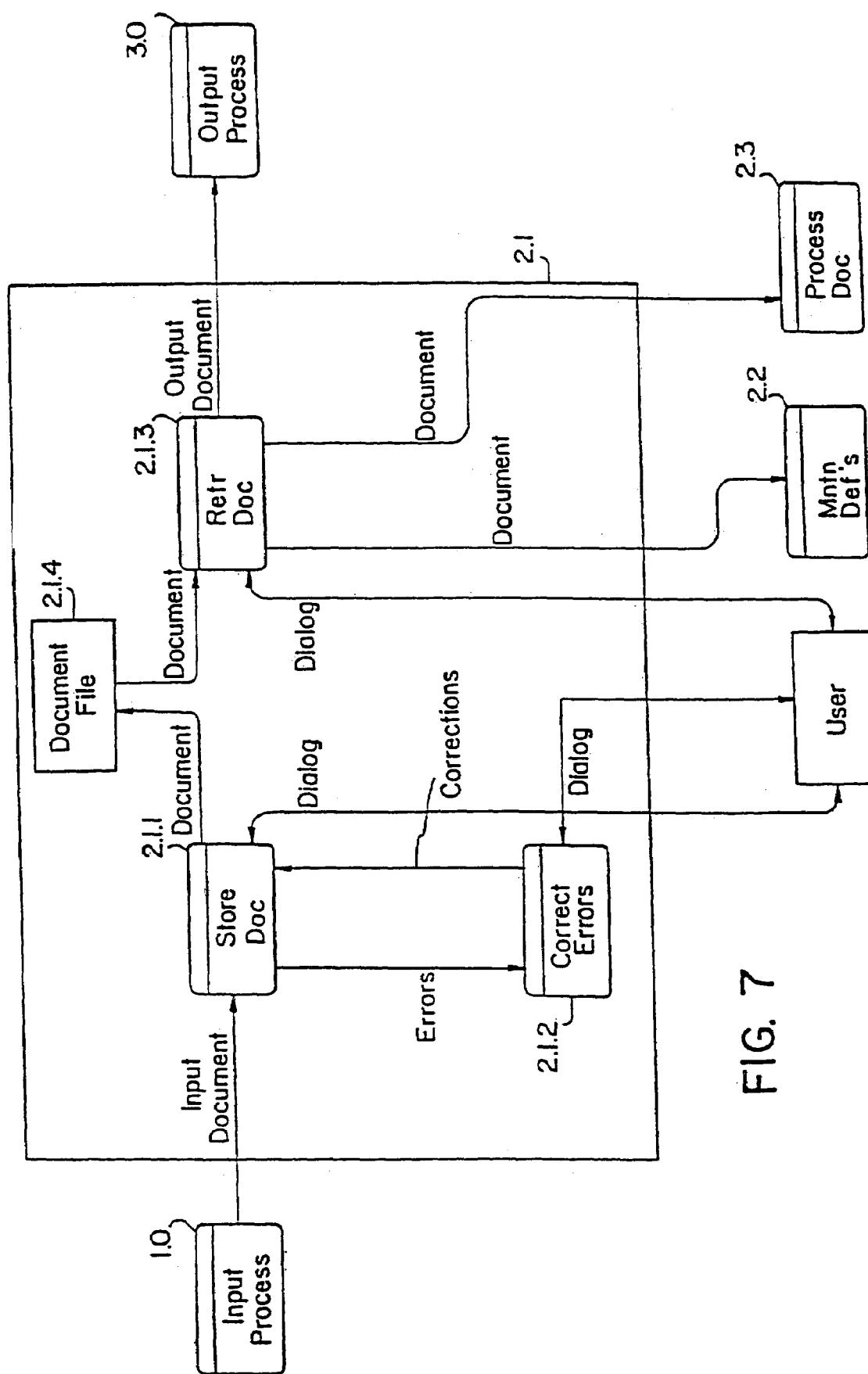


FIG. 7

U.S. Patent

Feb. 27, 2007

Sheet 9 of 15

US 7,184,162 B2

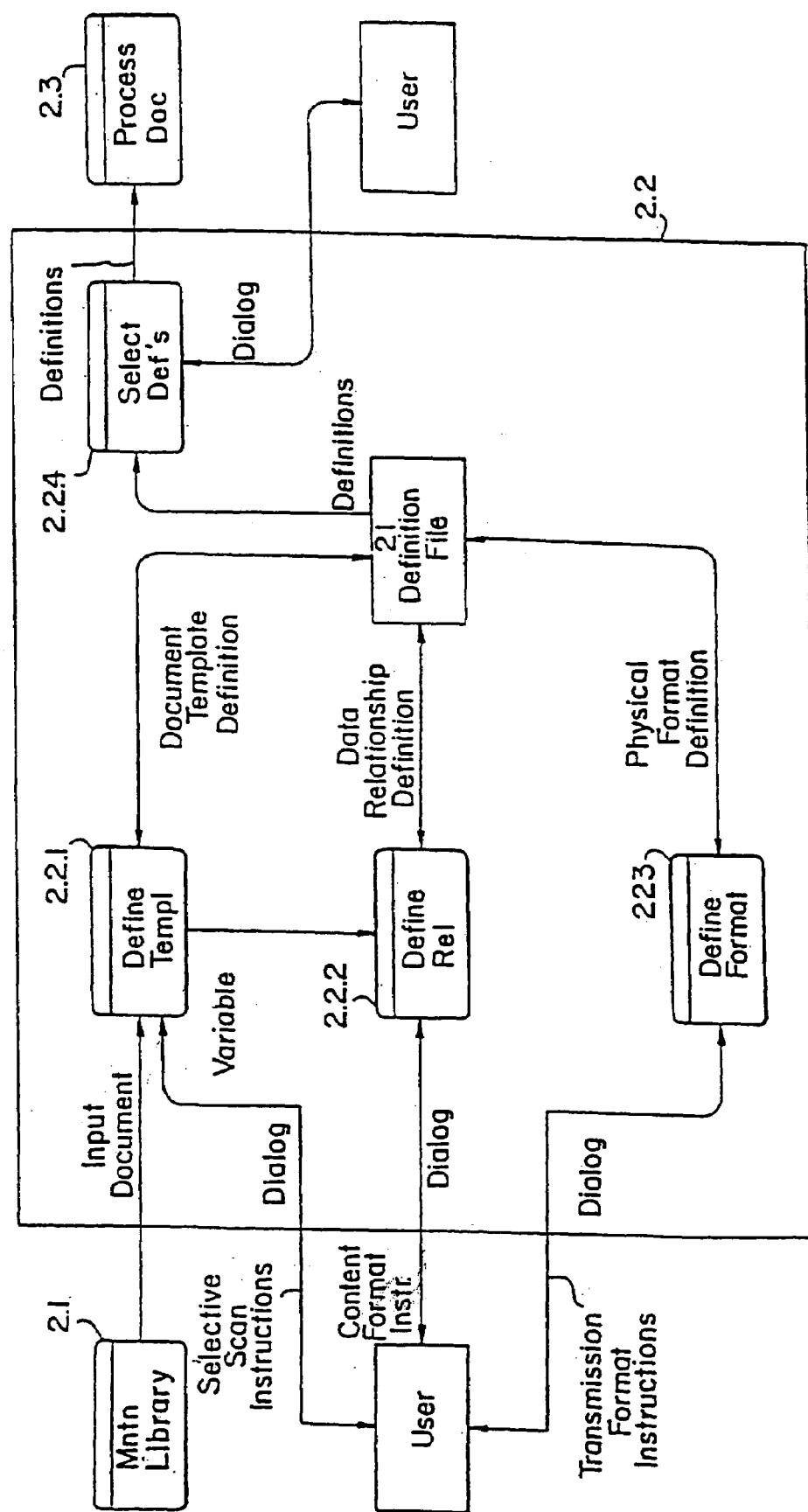


FIG. 8

U.S. Patent

Feb. 27, 2007

Sheet 10 of 15

US 7,184,162 B2

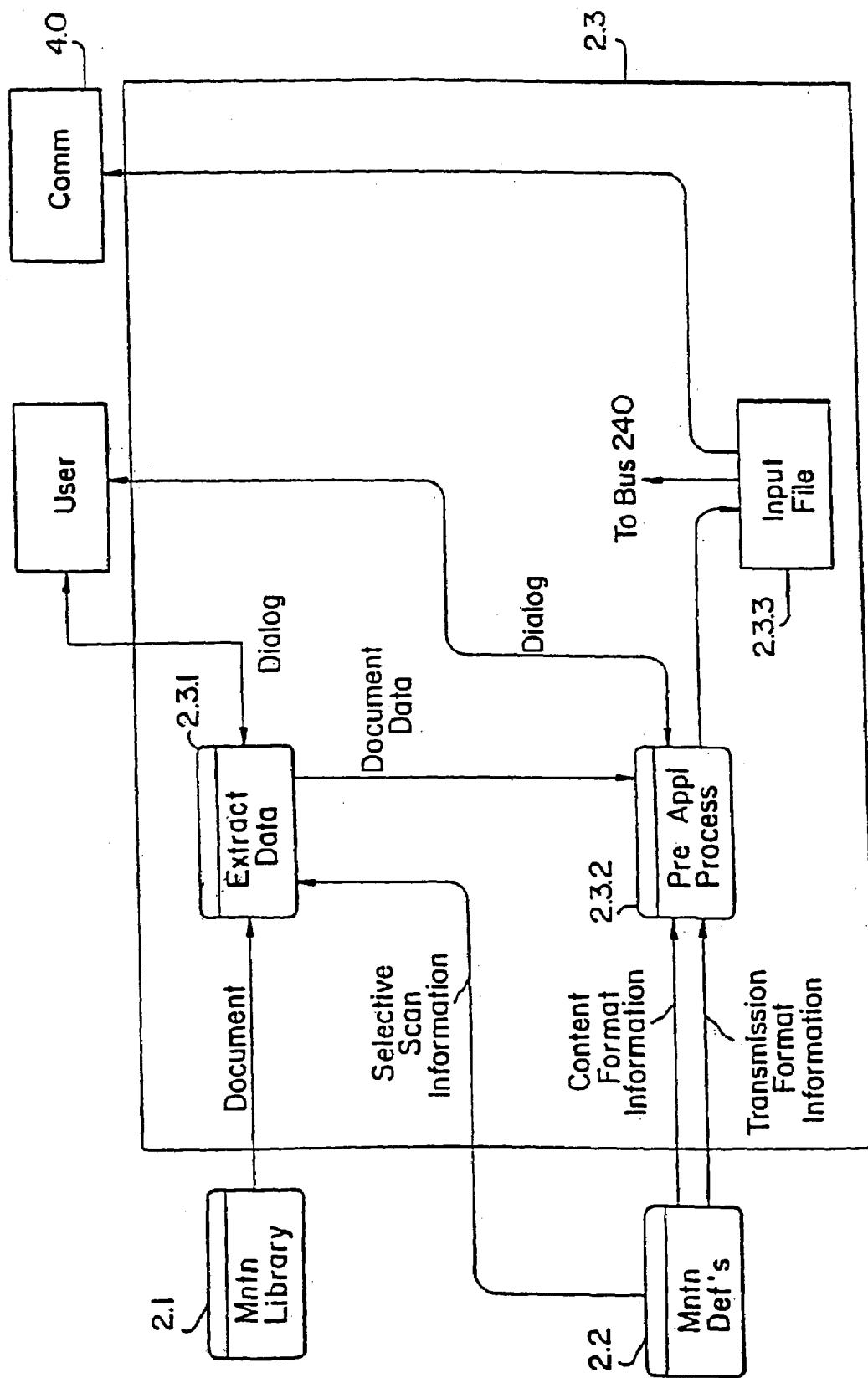


FIG. 9

U.S. Patent

Feb. 27, 2007

Sheet 11 of 15

US 7,184,162 B2

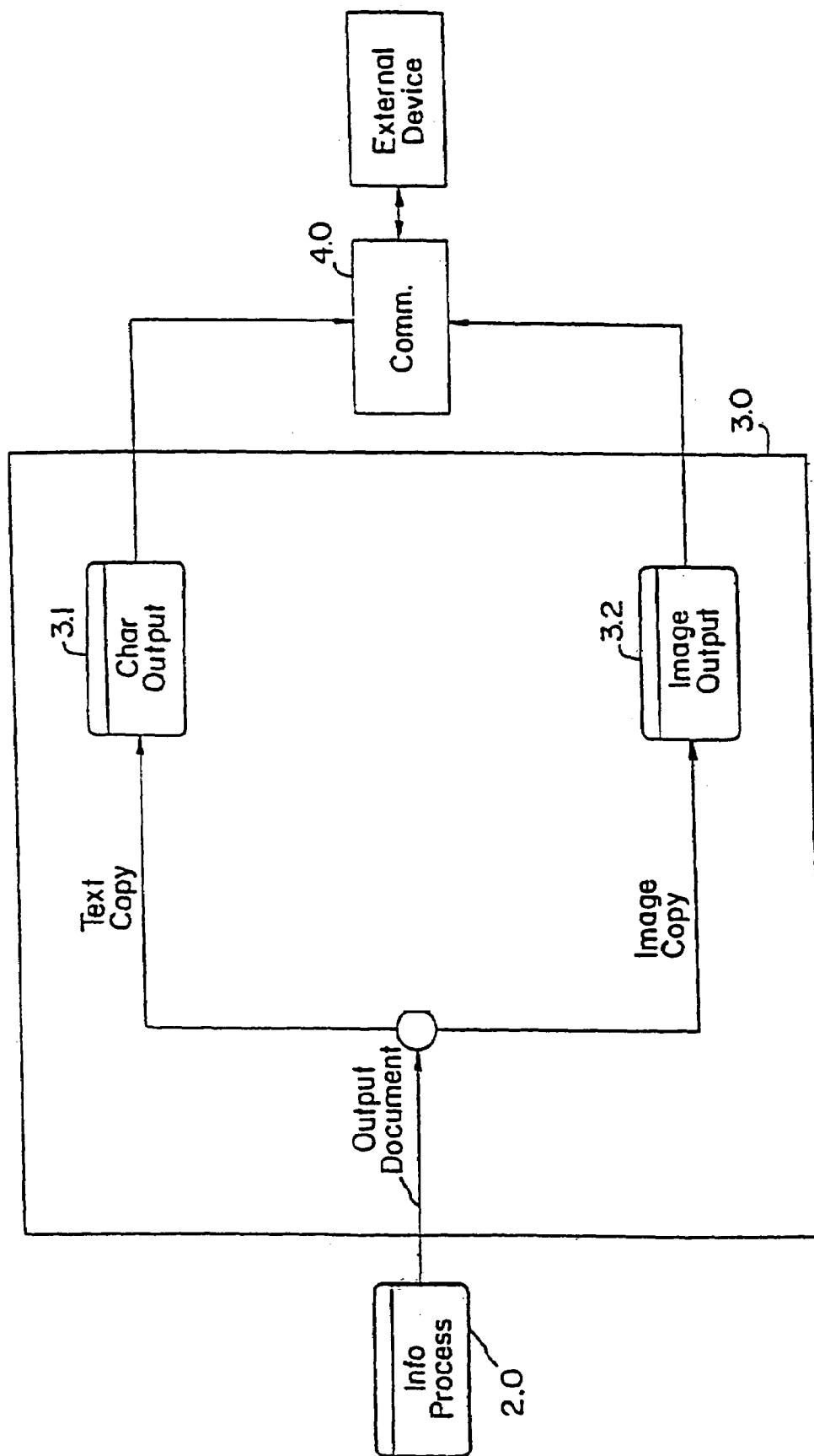


FIG. 10

U.S. Patent

Feb. 27, 2007

Sheet 12 of 15

US 7,184,162 B2

FIG. 11

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Heading 2	2
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
Other Credits	
New Balance	\$1025.91

U.S. Patent

Feb. 27, 2007

Sheet 13 of 15

US 7,184,162 B2

FIG. 12A

<u>Variable Name</u>	<u>Value</u>
Vendor	XYZ Corporation
Account Number	123456789
Statement Date	12/01/86
Payment Date	1/01/87
Previous Balance	\$1234.56
New Charges	\$789.01
Debits	
Finance Charges	\$2.34
Payments	\$1000.00
New Balance	\$1025.91

FIG. 12B

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

FIG. 12C

<u>Variable Name</u>	<u>Value</u>
Mail To	XYZ Corporation PO Box 567 Anywhere, NY 63130
Previous Balance	\$1234.56

U.S. Patent

Feb. 27, 2007

Sheet 14 of 15

US 7,184,162 B2

FIG. 13A

{ > 1 > 1 > 1 > "XYZ Corporation"
> 2 > 2 > 25 > + 123456789 >
> 3 > 2 > 1 > D12 / 01 / 86 >
> 4 > 2 > 11 > D12 / 15 / 86 >
> 5 > 2 > 21 > D01 / 01 / 87 >
> 6 > 10 > 25 > \$1234.56 >
> 7 > 11 > 25 > \$789.01 >
> 8 > 13 > 25 > \$2.34 >
> 9 > 14 > 25 > \$1000.00 >
> 10 > 16 > 25 > \$1025.91 >

FIG. 13B

{ > 1 > 1 > 1 > \$1234.56 >
> 2 > 2 > 1 > "XYZ Corporation"
> 3 > 3 > 1 > "PO Box 567"
> 4 > 4 > 1 > "Anywhere, NY 63130"

FIG. 13C

{ / 1 / 1 / \$1234.56 //
/ 2 / 2 / 1 / *XYZ Corporation*
/ 3 / 3 / 1 / *PO Box 567*
/ 4 / 4 / 1 / *Anywhere, NY 63130*

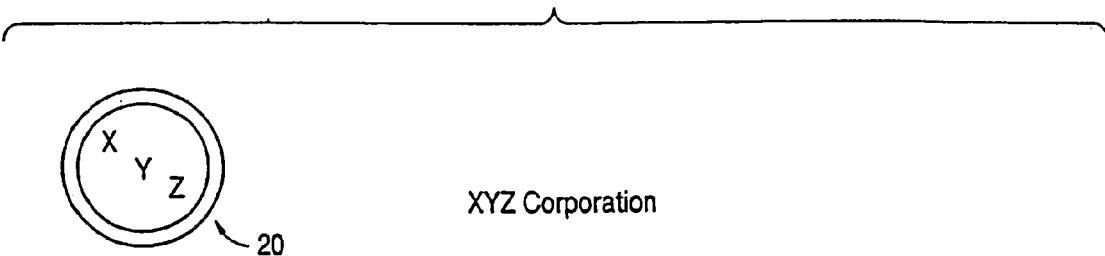
U.S. Patent

Feb. 27, 2007

Sheet 15 of 15

US 7,184,162 B2

FIG. 14



Customer

ABC Corporation
123 Sixteenth Street
Hometown, NJ 88981

Mail To:

△ XYZ Corporation
PO Box 567
Anywhere, NY 63130 △

Account Number 123456789

Previous Balance \$1234.56

Statement Date: 12/01/86

New Charges \$789.01

Payment Date: 1/01/87

Other Debits

Finance Charges (10%) \$2.34

Payments \$1000.00

WINTER SALE IN EFFECT THROUGHOUT JANUARY:

Other Credits

New Balance \$1025.91

30

US 7,184,162 B2

1

INFORMATION PROCESSING METHODOLOGY

This application is a continuation of application Ser. No. 10/704,484, filed Nov. 6, 2003, which is a continuation of application Ser. No. 09/458,162, filed Dec. 9, 1999, which is a continuation of application Ser. No. 09/044,159, filed Mar. 19, 1998 (now U.S. Pat. No. 6,094,505), which is a continuation of application Ser. No. 08/487,150, filed Jun. 7, 1995 (now U.S. Pat. No. 5,768,416), which is a divisional of Ser. No. 08/348,224, filed Nov. 28, 1994 (now U.S. Pat. No. 5,625,465), which is a continuation of Ser. No. 08/143,135, filed Oct. 29, 1993 (now U.S. Pat. No. 5,369,508), which is a continuation of Ser. No. 07/672,865, filed Mar. 20, 1991 (now U.S. Pat. No. 5,258,855).

BACKGROUND OF THE INVENTION

The invention is directed to a system for efficiently processing information originating from hard copy documents. More specifically, the invention is directed to a hard copy document to application program interface which minimizes the need to manually process hard copy documents.

In the past, information contained on hard copy documents was manually entered into a computer via the input controller of a particular computer. The original document was then filed away for future reference. Automatic input of data was limited to the input of Magnetic Ink Character Recognition (MICR) data and to Optical Character Recognition (OCR) data. This fixed-position data was forwarded directly to a dedicated computer application specifically designed to accommodate the input format. In more recent years, typewritten text has been mechanically inputted into a computer via a text file. Examples of this latter type of system are word processors and photo-typesetters.

These conventional systems have limitations which decrease the efficiency of processing information from a hard copy document. For example, the systems discussed above are limited in their application to MICR, OCR, or typewritten data. Parsing and processing data is limited to the particular requirements of the particular computer application which requires the input data. In addition, in these conventional systems, the actual hard copy document must be retained for future reference at great expense.

In a sophisticated computer network, different users may require different portions of the information contained on a hard copy document. For example, if the hard copy document is an invoice returned with payment of a bill, the accounting department may need all of the monetary information contained on the bill while the mailroom may need only customer address information, to update a customer's address. Therefore, there is a need for a system in which specific information from a hard copy document can be selectively distributed to various users.

Another problem with conventional systems is that users, even within the same company, may require that the information extracted from a hard copy document be transmitted to a particular application program in a specific transmission format. For example, one department in a company may use a particular application program which must receive information using a particular character as a delimiter and other departments may require the information in a different format using different delimiters.

Another problem, particularly for small businesses, is that current systems can not efficiently accommodate the inputting of information from a diversity of hard copy documents.

2

A large business which receives many forms in the same format can afford a system which inputs a high volume of information in that format into memory. For example, it is cost-effective for a bank which processes hundreds of thousands of checks a month to buy a dedicated machine which can read information off of checks having a rigidly defined, or fixed, format. However, as the diversity of forms received by a business increases relative to the number of forms that must be processed, it becomes less cost-effective to design a dedicated machine for processing each type of form format. This problem is particularly significant in small businesses which may, for example, receive fifty invoices a month, all in different, non-fixed, formats. It is frequently not cost-effective for a small business to design dedicated systems for inputting information in each of these various formats. This leaves a small business with no other practical alternative than to manually input the information off of each invoice each month.

SUMMARY OF THE INVENTION

It is an object of the invention, therefore, to provide an application program interface which allows a user to select specific portions of information extracted from a diversity of hard copy documents and allows the user to direct portions of this information to several different users in accordance with the needs of the particular user.

It is also an object of the invention to provide a cost-effective system for inputting hard copy documents which can accommodate hard copy documents in a diversity of formats.

It is another object of the invention to provide an application program interface which allows a user to put information, which is to be transmitted, into a particular transmission format, based upon the needs of the receiver of the information.

It is a further object of the invention to provide an application program interface which will allow the extraction, selection, formatting, routing, and storage of information from a hard copy document in a comprehensive manner such that the hard copy document itself need not be retained.

It is another object of the invention to provide a system which reduces the amount of manual labor required to process information originating from a hard copy document.

A further object of the invention is to reduce the time required to process information originating from a hard copy document so that a higher volume of transactions involving hard copy documents can be processed.

The invention provides an application program interface which inputs a diversity of hard copy documents using an automated digitizing unit and which stores information from the hard copy documents in a memory as stored document information. Portions of the stored document information are selected in accordance with content instructions which define portions of the stored document information required by a particular application unit. Selected stored document information is then formatted into the transmission format used by the particular application program based on transmission format instructions. The transmission formatted selected stored document information is then transmitted to the particular application program. The hard copy documents may contain textual information or image information or both.

The interface operates in three different modes.

In a first mode, the interface extracts all of the information from hard copy documents and stores this information in

US 7,184,162 B2

3

memory. Parsing of various portions of the extracted information is performed in accordance with content instructions.

In a second mode, the user operates interactively with the interface by use of a display and an input device, such as a mouse. In this second mode, a hard copy document is inputted and displayed on the display. The interface then prompts the user to identify the location of various information. For example, the interface can ask the user to identify the location of address information on the hard copy document. In response, the user positions the mouse to identify address information using a cursor. The identified information is then stored as address information in memory. Subsequently, the interface again prompts the user to identify other pieces of information, which are then stored in the appropriate locations in memory. This process proceeds until all of the information which is desired to be extracted off of the hard copy document is stored in memory.

In a third mode of operation, selected portions of information are extracted off of hard copy documents in accordance with predetermined location information which has been specified by the user. For example, the user can define a template which specifies the location of information on hard copy documents. Templates can be formed in conjunction with second mode operation. Alternatively, the user can instruct the interface to search hard copy documents for a particular character or symbol, located on the hard copy documents. The information desired to be extracted off of the hard copy documents is specified relative to the location of this character or symbol.

The interface can also prompt or receive from an applications program or another information processing system, required information, content instructions, and format instructions.

Other objects, features, and advantages of the invention will be apparent from the following detailed description of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be described in further detail below with reference to the accompanying drawings, in which:

FIG. 1 illustrates hardware for implementing a preferred embodiment of the instant invention;

FIG. 2 illustrates an example of a hard copy document containing information to be processed by the instant invention;

FIGS. 3A and 3B are enlarged views of the computer of FIG. 1 used to explain how the invention interactively prompts a user to identify information;

FIG. 4 is an overall data flow diagram for the FIG. 1 preferred embodiment;

FIG. 5 is a detailed input data flow diagram for the FIG. 1 preferred embodiment;

FIG. 6 is a detailed information processing data flow diagram for the FIG. 1 preferred embodiment;

FIG. 7 is a more detailed information processing data flow diagram for the maintain library module of FIG. 6;

FIG. 8 is a more detailed information processing data flow diagram for the maintain definitions module of FIG. 6;

FIG. 9 is a more detailed information processing data flow diagram for the process document module of FIG. 6;

FIG. 10 is a detailed output data flow diagram for the FIG. 1 preferred embodiment;

FIG. 11 lists data corresponding to the hard copy document of FIG. 2;

4

FIGS. 12A, 12B, and 12C illustrate examples of data which can be selected from the extracted data of FIG. 11 in accordance with content instructions;

FIGS. 13A, 13B, and 13C illustrate examples of the data of FIGS. 12A, 12B, and 12C formatted in accordance with various transmission format instructions to form input files; and

FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

15 Hardware

The invention provides an interface between information originating from a hard copy document and a computer application unit which uses the information. The computer application unit can be a particular computer application program or a device which is controlled in accordance with instructions or information from the hard copy document. The invention also allows storing a copy of the hard copy document in a memory and retrieving the copy of the hard copy document. By providing a comprehensive and integrated system which can accommodate almost all of the possible uses of information contained on a hard copy document, the instant invention allows for a paperless office.

The invention includes hardware and software necessary to extract, retrieve, and process information from the hard copy document. A copy of the actual image of the hard copy document is stored in memory. Textual information extracted from the hard copy document is also stored in memory. Textual information is information, such as alphanumeric characters, which is recognized on the hard copy document and which is stored in a form which corresponds to the particular recognized character. For example, the extracted characters can be stored in the ASCII format in an electronic memory.

40 The user can have all of the information extracted from the hard copy document and stored in memory. Alternatively, the interface can interactively prompt the user to identify specific pieces of information for storage. The interface can also extract specific pieces of information using a predefined template. The interface can also prompt or receive from another information processing system or an applications program desired information, content instructions, and format instructions.

45 The instant invention also provides for parsing information extracted from the hard copy document and for directing this parsed information to specific users or application programs as an input file.

The invention also permits the user to define the transmission format of the input file for a particular computer application unit.

50 FIG. 1 illustrates hardware for implementing a preferred embodiment of a hard copy document to application program interface according to the instant invention. The interface 200 processes information extracted off of hard copy document 100 and provides information to application units 270 in a form required by each particular application unit. The interface extracts information off of a hard copy document 100 utilizing a scanner 210. The scanner 210 can be any type of scanner which extracts information off of hard copy documents, for example, an Optical Reader.

55 The scanned information is stored in a scanner memory 220 or in main memory 250, as will be described in greater

US 7,184,162 B2

5

detail below. If main memory 250 or another memory is available to store the scanned information, then scanner memory 220 can be omitted.

The information from scanner memory 220 or main memory 250 is transmitted to computer 230. In the preferred embodiment, computer 230 includes a display 232, a keyboard 234, and a mouse 236. The display 232 displays an image of the hard copy document itself and/or information necessary to process the information extracted off of the hard copy document.

The computer 230 is used to select portions of the stored document information contained in memory in accordance with content instructions which define portions of the stored document information required by an application unit. These content instructions may be provided by the application program. Alternatively, the content instructions can be inputted via an input device such as a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

The computer 230 is also used to format selected stored document information into the transmission format used by an application unit based on transmission format instructions. The transmission format instructions may be provided by the application program. Alternatively, the transmission format instructions can be inputted via a keyboard, a touch screen, a mouse, a notepad, a voice recognition device, or the like.

Thus, the computer 230 is used to generate an input file for a particular application unit. The computer 230 is connected to scanner memory 220, main, or permanent, memory 250, a printer 260, and application units 270, via bus 240. Although FIG. 1 illustrates use of a bus to connect components together, it is understood that any routing or connecting link, implemented in hardware or software or both, can be employed instead of, or in addition to, a bus. Instructions to or in the computer 230 control the main memory 250, the printer 260, the application units 270, and the bus 240. Instructions to or in computer 230 can also control exchanges of information with scanner memory 220.

When the computer 230 generates an input file for a particular document, the computer 230 can send this input file directly to an application unit or can store this input file in the main memory 250 until required by an application unit. The main memory 250 may also optionally store a copy of the image information for the hard copy document and the textual information for the hard copy document. Thus, the image information and textual information from the hard copy document can be retrieved and printed out on printer 260. In addition, image and textual information stored in scanner memory 220 or in main memory 250 can be used to form additional input files at the time of input or at a later time, based on content instructions and transmission format instructions. Thus, the invention can, at the discretion of the user, eliminate the need to retain copies of hard copy documents, permitting a paperless office.

The application units 270 include particular application programs and devices which are controlled in accordance with information contained on hard copy document 100.

FIG. 2 illustrates an example of a hard copy document 100 which contains information to be processed by the instant invention. The document illustrated in FIG. 2 is a bill from XYZ Corporation to customer ABC Corporation. FIG. 2 is only an example of a type of document that can be processed by the instant invention.

In a first operational mode, the scanner 210 stores all of the information extracted off of hard copy document 100 in the scanner memory 220 or, alternatively, in main memory 250. The extracted information is stored in two forms. The

6

actual image of the hard copy document 100 is stored as image information in the scanner memory 220. In addition, the scanner memory 220 stores textual information recognized on the hard copy document 100 by, for example, employing standard character recognition software. In the preferred embodiment, the textual information is stored in ASCII format. The scanner memory 220 can be, for example, an electronic, magnetic, or optical memory.

FIG. 3A illustrates an enlarged view of the computer 230 of FIG. 1. This view will be used to describe a second mode of operation. In this second mode of operation, the hard copy document 100 is scanned and a copy of the document 100 is displayed on display 232 of computer 230, based on the contents of information temporarily stored in scanner memory 220. After the document is displayed on display 232, the computer 230 interactively prompts the user to identify the location of specific pieces of information on the hard copy document. In the FIG. 3A illustration, this prompt message is indicated as the message beginning with the arrow.

For example, the prompt message can ask the user to identify the location of account number information on the hard copy document. The user then uses an input device, such as keyboard 234 or mouse 236 or a touch screen, notepad, voice recognition device, or other input device to position a cursor on the display to identify the location of the information requested by the prompt message. For example, the cursor could be used to define a block (which could be highlighted) containing the requested information, followed by a mouse "enter" click. In this example, the user would move the mouse to identify the location of the account number information contained on the hard copy document 100. The computer 230 then stores the information which has been identified by the user as account number information in the appropriate address or subfile or as the appropriate variable or parameter in memory. The computer then prompts the user to identify the location of other information on the hard copy document, such as, statement date information. The process proceeds until all of the desired information has been stored into the appropriate locations in memory.

FIG. 3B illustrates a variation of the second mode for interactively prompting the user for information. In FIG. 3B, the display is split into two portions. A left-hand portion 232L displays the image of the hard copy document and a right-hand portion 232R displays the required application program information. For example, in FIG. 3B, portion 232R displays a spreadsheet used by an application program. While observing the split display, the user can input instructions to associate specific pieces of information on the hard copy document (for example, the vendor name indicated by the mouse arrow 232A) with particular subfiles in memory (for example, the vendor field next to which the cursor 232C appears), using a mouse or other input device(s) or both. The split display also allows the user to generate content format instructions while observing the information required for a particular application program on the right-hand portion.

These second modes of operation are efficient for small businesses which receive a small number of a wide variety of invoices, since the user does not necessarily have to store all of the information that appears on the hard copy document. A further advantage is that data input is quicker, easier, and more accurate than with previous keyboard methodology. In addition, by specifying the location on the hard copy document of information, the user may optionally create a template, to be described in further detail below, for each

US 7,184,162 B2

7

different type of invoice. This template is stored for future use when another hard copy document in the same format is received.

More specifically, instructions from computer 230 can direct the scanner 210 and scanner memory 220, and/or main memory 250, to scan and/or store only specific portions of hard copy document 100. After the interactive prompts required to obtain information for a desired application program, the unused information stored in scanner memory 220 or 250 can be erased. Further, scanning of a second identical document can be limited to only those portions of the document which contain needed information.

More specifically, in FIG. 2, the lines 10 drawn around certain portions of the document represent the areas which the user has previously identified as the portions of a document to be extracted by the scanner 210 and stored in scanner memory 220 and/or main memory 250. Since the logo 20 and the message 30 have not been identified as an area to be scanned and stored, these areas are not scanned and stored in subsequent documents. Since the user has previously associated each of the areas 10 with a specific subfile of information, e.g., the account number, the scanned information is stored in memory locations corresponding to that subfile.

Data Processing

FIGS. 4-10 illustrate the flow of data in the FIG. 1 preferred embodiment. FIG. 4 illustrates the overall data flow for the FIG. 1 preferred embodiment. The preferred embodiment includes an input process module 1.0, an information processing module 2.0, and an output processing module 3.0. The information processing module 2.0 is equipped to receive instructions from and transmit information to a user. The information processing module 2.0 can also transmit to and receive information from a remote external device through communication interface 4.0. Input process module 1.0 and output processing module 3.0 can also access communication interface 4.0. A module is implemented in hardware, software, or a combination of hardware and software. The specific implementation for a particular business application depends upon a variety of factors, for example, the relative costs of hardware and software implemented systems, the frequency with which a user will want to expand or modify the system, and the like.

FIG. 5 is a more detailed diagram of the input process module 1.0 of FIG. 4. The input process module 1.0 includes a character input module 1.1, an image input module 1.2, and, in the preferred embodiment, a character recognition device 1.3. The character input module inputs textual information, such as alphanumeric characters, from an input device such as keyboard 234. The image input module 1.2 inputs image information, for example, a digitized image of the actual appearance of hard copy document 100. Textual information can include textual input from an input device such as keyboard 234 and textual information extracted from the document by character recognition device 1.3. Both types of information comprise an input document which is transmitted to information processing module 2.0. In the FIG. 1 preferred embodiment, the processing performed by input process module 1.0 occurs in scanner memory 220, computer 230, and main memory 250.

FIG. 6 illustrates information processing data flow for the FIG. 1 preferred embodiment, that is, FIG. 6 illustrates data flow in the information processing module 2.0.

The information processing module 2.0 includes a maintain library module 2.1, to be described in further detail

8

below in conjunction with FIG. 7, a maintain definitions module 2.2, to be described in further detail below in conjunction with FIG. 8, and a process document module 2.3 to be described in further detail below in conjunction with FIG. 9.

The information processing module 2.0 is the module which coordinates and drives the entire system. In the preferred embodiment, the information processing module 2.0 is implemented primarily by computer 230.

FIG. 7 illustrates information processing data flow in the maintain library module 2.1. The maintain library module 2.1 maintains a library of image information, for example, a digitized image representing the actual appearance of the hard copy document, and textual information of the hard copy documents for reference during processing. This library can be incorporated within scanner memory 220, main memory 250, or another independent memory, for example, a RAM disk. The maintain library module 2.1 includes a store document module 2.1.1, a correct errors module 2.1.2, a retrieve document module 2.1.3, and a document file 2.1.4. These modules operate collectively to store, retrieve, and correct document information.

The store document module 2.1.1, prior to routing the document to the document file 2.1.4, may provide information on recognition errors which may have occurred while inputting the document. For example, the store document module 2.1.1 identifies that a character contained on hard copy document 100 was not recognized. The store document module 2.1.1 also optionally causes a copy of the document and its parsing to be displayed on the display 232 for confirmation by the user. The user may utilize this opportunity to identify any errors in the displayed document and, in conjunction with the correct errors module 2.1.2, to revise the document's parsing, if necessary, prior to storage of the document in memory. The module 2.1.1 also provides a facility for the user to name a particular hard copy document for cataloging, storage, and retrieval purposes. After the document is named, the store document module 2.1.1 stores copies of the document in the document file 2.1.4.

The correct errors module 2.1.2 processes instructions from the user to correct errors identified by the store document module 2.1.1 and errors that have been spotted by the user during the confirmation process.

The retrieve document module 2.1.3 permits the user to retrieve a copy of a document previously stored in the document file 2.1.4. As described above, long-term storage is provided by main memory 250, if necessary.

FIG. 8 illustrates a more detailed information processing data flow diagram for the maintain definitions module 2.2 of FIG. 6. The maintain definitions module 2.2 allows the user to define system and document parameters and maintains the definitions of these system and document parameters. The maintain definitions module 2.2 includes a define template module 2.2.1 which allows the user to specify the location of information on the document. This information provided by the user defines a template which is used to extract information off the document and to associate the extracted information with a particular variable or subfile. These templates are illustrated by boxes 10 in the FIG. 2 example of a hard copy document. The maintain definitions module 2.2 can also access templates previously defined by the user and stored in main memory 250. Templates can also be provided as part of software packages developed by program developers.

The maintain definitions module 2.2 also includes a define relationships module 2.2.2. The define relationships module 2.2.2 allows the user to define data relationships, or logical

US 7,184,162 B2

9

relationships, between pieces of information extracted from the hard copy document. These pieces of information are then used to generate an input file for a selected computer application unit. The user defines these relationships by content instructions. Alternatively, content instructions to define relationships can be provided by application software. If the user provides these content instructions, the content instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. Examples of content instructions, data, and logical relationships will be described in further detail in conjunction with FIGS. 11 and 12A, 12B, and 12C.

The maintain definitions module 2.2 also includes a define format module 2.2.3. The define format module 2.2.3 allows the user to define transmission formats for an input file which is then transmitted to a selected computer application unit. Selection of the transmission format of the input file is accomplished by the user through use of transmission format instructions. Alternatively, the applications software itself can generate its own transmission format instructions. When the user must specify transmission format instructions, the transmission format instructions are inputted via keyboard 234 or via another input device such as a notepad, a voice recognition device, or the like. A further description of various transmission formats will be provided below in conjunction with FIGS. 12A, 12B, 12C, 13A, 13B, and 13C.

A select definitions module 2.2.4 is also included in the maintain definitions module 2.2. The select definitions module 2.2.4 allows the user to store and select a set of definitions to be used for processing the document. The definitions identify pieces of information on the document by, for example, absolute location, variable location, or relative location, or by proximity to key words and/or symbols. These definitions are described in further detail below by way of an illustrative example.

FIG. 9 illustrates a more detailed information processing data flow diagram for the process document module 2.3. The process document module 2.3 processes the document after the document has been stored in the system. The process document module 2.3 gathers the appropriate information which has been stored, and creates input file(s) 2.3.3 for the selected application unit. The process document module 2.3 then transmits the input file(s) via bus 240 and/or communication interface 4.0 to an application unit 270, an output device such as printer 260, or to main memory 250.

The process document module 2.3 includes an extract data module 2.3.1. This module extracts data off of the document in accordance with the user's instructions, for example, the user-defined template, or through the interactive mode.

The process document module 2.3 also includes a preapplication process module 2.3.2 which gathers and associates information extracted from the document in accordance with content instructions. This module prompts the user for any additional information required to satisfy the relationships defined by the content instructions. The preapplication process module 2.3.2 also places the selected information into the transmission format defined by the transmission format instructions.

The preapplication process module 2.3.2 also generates the input file 2.3.3 for the selected application in accordance with the appropriate instructions. The input file 2.3.3 is then transmitted to bus 240 and/or communication interface 4.0 for transmission to a particular application unit 270.

FIG. 10 illustrates a detailed output data flow diagram for output module 3.0. Output module 3.0 outputs a textual and/or image copy of the document. In the FIG. 1 preferred

10

embodiment, output module 3.0 is implemented by printer 260, associated software, and associated interface circuitry.

Operation

5

Examples of operation of a preferred embodiment will now be described.

The user enters the system by providing instructions to the information processing module 2.0. The user then instructs the information processing module 2.0 to conduct maintain library processing, maintain definitions processing, or process document processing.

If the user selects maintain library processing, the user then provides instructions to maintain or modify the document library through the maintain library module 2.1. For example, the user can direct the inputting and storage of a hard copy document 100 or can retrieve and output a document. The user requests inputting of a document through the store document module 2.1.1. The system then prompts the user to specify a storage location for the inputted document. The document is then read-in by the input process module 1.0. A textual copy and/or an image copy are stored into the document file 2.1.4. Errors which have occurred during inputting are identified and corrected by the correct errors module 2.1.2 and the user. The corrections are reflected in the document information stored in document file 2.1.4.

The retrieve document module 2.1.3 is used to retrieve and output a document. The system prompts the user to specify the storage location of a document and the type of document copy, for example, a textual or an image copy, to be outputted. The document is then outputted by the output process module 3.0.

If the user initially selected maintain definitions processing, the user would instruct the system to maintain and/or modify parameter definitions through the maintain definitions module 2.2. For example, the user can define and maintain a document template for extracting selected portions of information off of the hard copy document. The user can use the template to extract selected portions of information off of the hard copy document when the document is originally inputted, or alternatively, the user can use the template to identify selected portions of information for extraction off of an image copy of the document. In creating the template, the user identifies pieces of information on the document to be extracted and assigns a variable name, or subfile, to each piece of data.

The location of data to be extracted can be defined in a number of ways other than by use of a template. For example, the user can designate the absolute location of information on the document with respect to a grid overlaid on the document, e.g., always on line 3, starting in column 1. The user can also identify information by specifying the relative location of information to be extracted, e.g., always two lines below the piece of data named "salutation", starting in column 3. The user can also specify the location of information to be extracted by variable location specification. For example, if the hard copy document is a letter, the module would conduct a key word search for the term "Dear Sir:". Wherever this term "Dear Sir:" is located, this piece of data would be associated with the variable specified by the user, for example, the variable "salutation." In addition, a defined set of conventional symbols can be used to signify certain recurring data items for the convenience of users of the instant invention. For example, a "@" symbol can be used to delineate the vendor name as follows:

US 7,184,162 B2

11

“@XYZ Corporation@”. Other examples of the use of symbols to delineate information will be described with reference to FIG. 14.

The maintains definition module 2.2 is also used to maintain data relationships in accordance with content instructions and to maintain input file formats in accordance with transmission format instructions. Relationships are defined and maintained between pieces of data, specified by, for example, the names of variables, through the define relationships module 2.2.2. The names of pieces of data on the document are retrieved by, for example, the define template module 2.2.1, and are passed to the define relationships module 2.2.2. The user may then provide any additional pieces of data needed to generate an input file for a particular application program or unit, such as an input file line number. The user, the applications software, and/or instructions previously stored in memory then establishes the contents of the input file by defining relationships between pieces of data using content instructions. Specific examples of content instructions will be discussed below in conjunction with FIGS. 11, 12A, 12B, 12C, 13C, 13A, 13B, and 13C.

The user and/or the applications software defines and maintains the transmission format of the input file to be used by a particular application program or unit through the define format module 2.2.3 in accordance with transmission format instructions. This is accomplished by defining the parameters to be used by the preapplication process module 2.3.2 in generating an input file. Parameters which would typically be required to generate an input file would include the character type, e.g., text or pixel; delimiters used between pieces of data, e.g., a slash or a semicolon; end of line characters, e.g., a carriage return or a line feed; and end of file characters. Examples of transmission formats will be described in further detail below in conjunction with FIGS. 11, 12A, 12B, 12C, 13A, 13B and 13C.

If the user initially selected process document processing, the interface will then proceed to process the document through use of the process document module 2.3. For example, the user can extract specific portions of data from an image copy of a document, can generate an input file for transmission to an application program, or can directly process information interactively with an application program.

If the user desires to extract specific portions of data from an image copy of a hard copy document which has already been stored in memory, the user uses the extract data module 2.3.1 to identify a document to be processed. The document is then retrieved by the retrieve document module 2.1.3 and passed to the extract data module 2.3.1. The user can also select parameter definitions through the select definitions module 2.2.4.

The selected document template or parameter definition is passed to the extract data module 2.3.1. The extract data module 2.3.1 extracts pieces of data from the image copy of the document, as defined by the document template definition or the parameter definitions or both. This document data is then passed to preapplication process module 2.3.2.

The interface generates input file(s) 2.3.3 by use of the preapplication process module 2.3.2. The selected data relationship definition, as defined by the content instructions, and the selected record format definitions, as defined by the transmission format instructions, are passed to the preapplication process module 2.3.2. The preapplication process module 2.3.2 assembles the input file in accordance with the content instructions. The preapplication process module 2.3.2 also prompts the user for any additional pieces of data

12

which need to be provided by the user. The input file is converted to the desired transmission format in accordance with the transmission format instructions. This physically formatted data is then stored in the input file 2.3.3.

5 The user can also use an application program to process information by loading the particular application program into the computer 230 rather than by sending the input file to a remote application unit 270.

10 An illustrative example of the processing described above will now be described.

The user inputs instructions via keyboard 234 or another input device which indicate that the user desires to input and store a document. The computer 230 then prompts the user for the name of the document. In this example, the user desires to input the document of FIG. 2 and therefore names the document “XYZ Corp. Bill Dec. 1, 1986.” The computer then prompts the user to feed the hard copy document 100 into the scanner 210. The image of the hard copy document is displayed on display 232. The computer then prompts the user to identify the account number on the document. By use of the mouse 236 or other input device to position a cursor on the display, the user indicates the location of the account number. The account number is then read-in to a subfile named “Account Number.” This process proceeds until all of the desired information has been read-in and stored.

15 In this particular example, no errors were encountered while inputting the document. The user then directs that the document be stored for future reference in a document file.

20 Some time later, the user desires to retrieve and output the document and to generate input files based on information from the document. The computer 230 prompts the user for the name of the document and the type of output. The user responds with “XYZ Corp. Bill Dec. 01, 1986” for a printed textual copy. The document is then retrieved from the document file and passed to the printer 260 for printing.

25 In order to generate an input file for a specific application program, the user selects the option to define a document template for use when each month's XYZ Corporation bill arrives. Accordingly, the user instructs the system to display a copy of an XYZ Corporation bill on the display 232. The user then identifies pieces of data by absolute locations. That is, the user assigns specific names to information located at specific portions of the document. In this example, the user would input the following information:

30 Vendor-text, line 1, one line, column 1, 80 characters;
Account number-numeric, line 6, one line, column 25, 9 characters;

35 Statement date-date, line 9, one line, column 25, 8, characters;

Payment date-date, line 11, one line, column 25, 8 characters;

40 Previous balance-currency, line 7, one line, column 75, 9 characters;

New charges-currency, line 8, one line, column 75, 9 characters;

Other debits-currency, line 10, one line, column 75, 9 characters;

50 Finance charges-currency, line 12, one line, column 75, 9 characters;

Payments-currency, line 13, one line, column 75, 9 characters;

Other credits-currency, line 14, one line, column 75, 9 characters;

55 New balance-currency, line 15, one line, column 75, 9 characters.

US 7,184,162 B2

13

The user also identifies data with variable locations. In this particular example, a variable location is specified as follows:

Heading 2-line, value="Mail To:"

The identification of Heading 2 as line information means that the system will search for occurrences of the character string "Mail To:" and assign the line number which contains this character string to Heading 2.

The user also identifies data by relative locations. In this example, the user identifies the following relative location:

Mail To-text, Heading 2+1, 3 lines, column 60, 25 characters per line.

The instructions above instruct the system to assign the textual information beginning on one line after Heading 2 and continuing for 3 lines, in column 60, to the Mail To subfile.

As an alternative to inputting the actual line, column, and character numbers, the user can identify desired portions of the document by blocking, or highlighting, the desired portions using the mouse or other input device. In this case, the computer converts the highlighted portions into corresponding line, column, and character numbers.

FIG. 11 lists data corresponding to the hard copy document of FIG. 2 and the associated variable or subfile names.

Next, the user desires to define data relationships in accordance with content instructions. Examples of the type of contents which can be specified by a user are illustrated in FIGS. 12A, 12B, and 12C.

In this particular example, three separate departments of ABC Corporation require information from the XYZ Corporation bill. The first department requires vendor, account number, statement date, payment date, previous balance, new charges, debits, finance charges, payments, and new balance information. The second and third departments require mail to information and previous balance information. Each of these departments have their own application program which utilizes this information.

The user employs content instructions to designate how pieces of information, which have been extracted off of hard copy document 100, are directed to particular departments, that is, particular application programs. FIG. 12A illustrates the contents of the information to be transmitted to the first department. FIG. 12B illustrates the information to be transmitted to the second department. FIG. 12C illustrates the information to be transmitted to the third department. The content instructions, therefore, parse the information shown in FIG. 11 to various application programs, as shown by FIGS. 12A, 12B, and 12C. Content instructions can also be used to identify additional pieces of data which are required for the input files of the particular application programs. In this particular example, the specific application programs from the three departments all require numeric record number information, numeric horizontal position information, numeric vertical position information, and date received information. The horizontal and vertical position information is used by the application program to specify the location of the received information on a spreadsheet application program, in this example. The user may know in advance the content format required by each application program, that is, in this example, the location and type of information specified on the spreadsheet. The user may also employ the split display mode described with reference to FIG. 3B to generate content format instructions.

14

Using the content instructions, the user establishes the following contents for the input file corresponding to FIG. 12A:

5 Record number, horizontal position, vertical position, vendor;

Record number, horizontal position, vertical position, account number;

Record number, horizontal position, vertical position, statement date;

10 Record number, horizontal position, vertical position, date received;

Record number, horizontal position, vertical position, payment date;

Record number, horizontal position, vertical position, previous balance;

Record number, horizontal position, vertical position, new charges;

Record number, horizontal position, vertical position, finance charges;

20 Record number, horizontal position, vertical position, payments;

Record number, horizontal position, vertical position, new balance.

Next, transmission format instructions are employed to 25 define the transmission format of the input file for a specific application program or unit. FIG. 13A illustrates the transmission input file corresponding to FIG. 12A. FIG. 13B illustrates the transmission input file corresponding to FIG. 12B. FIG. 13C illustrates the transmission input file corresponding to FIG. 12C. A comparison of FIGS. 12B and 12C reveals that FIGS. 12B and 12C have the same contents. However, the information illustrated in FIG. 12B is being sent to a different application program than the information in FIG. 12C. These application programs require different 30 transmission input formats, as illustrated in FIG. 13B and 13C. More specifically, the application program that receives the input file illustrated in FIG. 13B uses the greater than sign as a delimiter whereas the application program which receives the transmission input file shown in FIG. 40 13C uses a back-slash as the delimiter.

After the contents and the transmission format for the input file have been, defined, and any additional information has been inputted, the input file is assembled and transmitted to the particular application program.

45 FIG. 14 illustrates another example of a hard copy document containing information to be processed by the instant invention. The hard copy document illustrated in FIG. 14 is first scanned and information from the hard copy document is stored into a memory. The interface 200 then identifies portions of the hard copy document corresponding to various variables by recognizing a defined set of symbols. In the FIG. 14 example, triangles delineate the mailing address, circles delineate the statement date, and squares delineate the new charges. Information from these portions of the hard 50 copy document is stored in the corresponding memory locations or subfiles for each variable. The same set of symbols can be used to identify the same information from one document to the next. Thus, even if the physical formats of documents are not fixed from one document to the next, a diversity of hard copy documents can be processed without 55 manually inputting data by recognition of the defined symbols.

60 Examples of readily available application programs are Quicken and Lotus 1,2,3 both of which are widely utilized in the business community. Quicken, for example is an easy-to-utilize program for writing checks and preparing business records. Payee, amount and address information

US 7,184,162 B2

15

may readily be transmitted from scanner memory 220 and/or main memory 250 to the Quicken application program for check writing functions and ledger keeping purposes. Lotus is a well known spreadsheet program which may process data input into specified cells once this data is placed in conventional Lotus format.

Thus, the instant invention provides an integrated and comprehensive system for handling information from a hard copy document, thus permitting a paperless office. In addition, the invention permits data, extracted off of a hard copy document, to be easily manipulated into various logical and transmission formats required by a particular application unit. The invention also provides a low cost system for inputting information from a wide variety of hard copy documents into a memory.

The foregoing description has been set forth merely to illustrate preferred embodiments of the invention and is not intended to be limiting. Modifications are possible without departing from the scope of the invention.

For example, letters, checks, forms, pictures, reports, music scores, film, and other types of hard copy documents can be processed by the invention for accounts payable/receivable accounting, inventory control, record keeping, budgeting, data base management, music transcription, forms processing, computerized art, survey and questionnaire processing, statistical data analysis, correspondence processing and other applications.

Other automated digitizing units can be used in addition to or as an alternative to use of the scanner 210 as an input unit. Any electrical, magnetic, or optical device which extracts information off of a hard copy document, thereby eliminating the need to manually input significant amounts of information from the hard copy document is suitable for use as an automated digitizing unit. In addition, information can be input by user responses and digital and analog signals generated from various devices, and from computer files from other computer systems. Suitable hardware for inputting data includes a keyboard, a light pen, a mouse, a touch screen, a laser scanner, a microphone, a tablet, a disk drive, a magnetic tape drive, and a modem.

The interface 200 can also output information in forms other than a hard copy of textual or image information. For example, the interface 200 can output system responses, computer files, and digital and analog signals for transmission to other computer systems or to control systems. Suitable hardware for outputting information includes a disk drive, a magnetic tape drive, a cathode ray tube, a plasma screen, a printer, a plotter, a film developer, an amplifier, and a modem.

Since modifications of the described embodiments incorporating the spirit and substance of the invention may occur to persons skilled in the art, the scope of the invention should be limited solely with respect to the appended claims and equivalents.

What is claimed is:

1. A method of doing business utilizing a computer system configurable to automatically extract information according to content instructions from an electronic document template file displayed on a first computer to a plurality of application programs running on at least one other computer and transmit said information according to customizable transmission format instructions to at least one field requiring said transmitted information in at least one application program running on said at least one other computer to enable a commercial transaction between a buyer and a seller.

2. The method as recited in claim 1 wherein the transaction involves a physical good.

16

3. The method as recited in claim 1 wherein the transaction involves a service.

4. The method as recited in claim 1 wherein the transaction involves copyrighted material.

5. The method as recited in claim 1 in which said first computer and at least one other computer are part of a distributed computing system.

6. The method as recited in claim 1 wherein said automatic extracting step automatically parses said information to said at least one field in said application program.

7. The method as recited in claim 1 wherein content instructions are automatically generated during said automatically extracting step.

8. The method as recited in claim 7 wherein said content instructions are automatically generated using a mouse function.

9. The method as recited in claim 7, wherein said content instructions are automatically generated using speech recognition.

10. The method as recited in claim 7 wherein said content instructions are automatically generated based on interaction by the user with at least one of said first and at least one other computer.

11. The method as recited in claim 1 wherein said extracted information comprises billing information.

12. The method as recited in claim 1 wherein said extracted information comprises financial information.

13. The method as recited in claim 1 wherein said extracted information comprises personal information.

14. The method as recited in claim 1 wherein said extracted information comprises customer information.

15. A method of doing business utilizing a computer system configurable to automatically extract information according to content instructions from an electronic document template file displayed on a first computer to a plurality of application programs running on at least one other computer and to transmit said information according to format instructions to at least one field requiring said transmitted information in at least one application program running on at least one other computer to enable a commercial transaction between a buyer and a seller.

16. The method as recited in claim 15 wherein the transaction involves a physical good.

17. The method as recited in claim 15 wherein the transaction involves a service.

18. The method as recited in claim 15 wherein the transaction involves copyrighted material.

19. The method as recited in claim 15 in which said first computer and at least one other computer are part of a distributed computing system.

20. The method as recited in claim 15 wherein said automatic extracting step automatically parses said information to said at least one field in said application program.

21. The method as recited in claim 15 wherein content instructions are automatically generated during said automatically extracting step.

22. The method as recited in claim 21 wherein said content instructions are automatically generated using a mouse function.

23. The method as recited in claim 21, wherein said content instructions are automatically generated using speech recognition.

24. The method as recited in claim 21 wherein said content instructions are automatically generated based on interaction by the user with at least one of said first and at least one other computer.

US 7,184,162 B2

17

25. The method as recited in claim 15 wherein said extracted information comprises billing information.

26. The method as recited in claim 15 wherein said extracted information comprises financial information.

27. The method as recited in claim 15 wherein said extracted information comprises personal information.

28. The method as recited in claim 15 wherein said extracted information comprises customer information.

29. A method of doing business by automatically extracting information from an electronic document template file displayed on a first computer and transmitting said information to at least one of a plurality of application programs running on at least one other second computer according to customizable transmission format instructions, said at least one other application program requiring said transmitted information to enable a commercial transaction between a buyer and a seller.

30. The method as recited in claim 29 wherein the transaction involves a physical good.

31. The method as recited in claim 29 wherein the transaction involves a service.

32. The method as recited in claim 29 wherein the transaction involves copyrighted material.

33. The method as recited in claim 29 in which said first computer and at least one other computer are part of a distributed computing system.

34. The method as recited in claim 29 wherein said automatic extracting step automatically parses said information to said at least one field in said application program.

35. The method as recited in claim 29 wherein content instructions are automatically generated during said automatically extracting step.

36. The method as recited in claim 35 wherein said content instructions are automatically generated using a mouse function.

37. The method as recited in claim 35, wherein said content instructions are automatically generated using speech recognition.

38. The method as recited in claim 35 wherein said content instructions are automatically generated based on interaction by the user with at least one of said first and at least one other computer.

39. The method as recited in claim 29 wherein said extracted information comprises billing information.

40. The method as recited in claim 29 wherein said extracted information comprises financial information.

41. The method as recited in claim 29 wherein said extracted information comprises personal information.

18

42. The method as recited in claim 29 wherein said extracted information comprises customer information.

43. A method of doing business by extracting information from a file displayed on a first computer, said file configurable to transmit information to a plurality of application programs running on at least one other computer and transmitting said information according to format instructions to at least one application program on said at least one other computer to enable a commercial transaction between a buyer and a seller.

44. The method as recited in claim 43 wherein the transaction involves a physical good.

45. The method as recited in claim 43 wherein the transaction involves a service.

46. The method as recited in claim 43 wherein the transaction involves copyrighted material.

47. The method as recited in claim 43 in which said first computer and at least one other computer are part of a distributed computing system.

48. The method as recited in claim 43 wherein said automatic extracting step automatically parses said information to said at least one field in said application program.

49. The method as recited in claim 43 wherein content instructions are automatically generated during said automatically extracting step.

50. The method as recited in claim 49 wherein said content instructions are automatically generated using a mouse function.

51. The method as recited in claim 49, wherein said content instructions are automatically generated using speech recognition.

52. The method as recited in claim 49 wherein said content instructions are automatically generated based on interaction by the user with at least one of said first and at least one other computer.

53. The method as recited in claim 43 wherein said extracted information comprises billing information.

54. The method as recited in claim 43 wherein said extracted information comprises financial information.

55. The method as recited in claim 43 wherein said extracted information comprises personal information.

56. The method as recited in claim 43 wherein said extracted information comprises customer information.

* * * * *